

# THIS WEEK

## EDITORIALS

**CONSERVATION** Vulture-killing drug gains foothold in Europe **p.404**

**WORLD VIEW** Four decades of bird study all at sea **p.405**



**RHINOS** How legal horn trade could drive away poachers **p.406**

## A ripe time for gaining ground

*After three years of heated debate, the advocates and critics of gain-of-function research must work to agree on how best to regulate the work.*

Late last week, the White House Office of Science and Technology Policy and the US Department of Health and Human Services announced an immediate pause in all new government funding for gain-of-function (GOF) research — experiments to boost the transmissibility, virulence or host range of pathogens — on influenzas, Middle East respiratory syndrome and severe acute respiratory syndrome.

The pause is to allow time to develop a new policy on how such work should be conducted and regulated. The policy will be informed by a full assessment of the risks and benefits of the work, and by how these compare with those for safer alternatives. In charge of that assessment is the National Science Advisory Board for Biosecurity (NSABB), which is meeting this week for the first time in two years (see page 411). The board will also advise on the policy's content. In parallel, the National Research Council (NRC) of the National Academies will convene a scientific conference on the issues surrounding GOF research, including its risks and benefits. It will also review the NSABB's draft recommendations for the new policy.

Controversy over GOF research was first sparked in late 2011 when the NSABB attempted to stop the publication of the full results of two studies in which the H5N1 avian flu virus had been engineered to become transmissible in mammals. The board and others were worried that information in the papers could help terrorists or other malevolent individuals to develop a bioweapon. Those concerns were finally overruled, and on 24 September, the United States adopted new rules on what is known as dual-use research — work that could be misapplied to do harm — on 15 pathogens or toxins.

A wider concern raised at the time — which has since shifted to front and centre — was the risk of a pathogen that had been engineered to become more dangerous escaping from the lab. In February 2012, the US Department of Health and Human Services added another layer of review for grant proposals involving GOF research, but only for H5N1; this was extended to H7N9 in August 2013. The research community became deeply polarized over the issues surrounding GOF work. Some vaunted the benefits of such research for pandemic preparedness and down-played biosafety and biosecurity risks, whereas others argued that the experiments should not be done because the risks far outweighed the benefits. To allow time for debate, GOF researchers agreed to put their research on hold, resuming work a year later after deciding that enough time had passed.

The decision to implement another moratorium — and to broaden it to pathogens other than the H5N1 and H7N9 flu viruses — is a belated acknowledgement that the issue of how to handle GOF research is far from resolved. And the revelations over the past few months of serious violations and accidents at some of the leading biosafety containment labs in the United States has burst the hubris that some scientists, and their institutions, have in their perceived ability to work safely with dangerous pathogens. The US administration cited these revelations as one reason for the latest review; behind-the-scenes

lobbying by critics of GOF research also played a part.

The climate for constructive discussion is now perhaps better than it was: although opinions remain sharply divided, each side now seems to be listening more to the other. In July, almost 300 scientists and policy experts signed up to the 'Cambridge Consensus', which criticized the lack of a proper risk-benefit assessment of the research, and called for exactly what the US government has now agreed to do. More than

***"The climate for constructive discussion is now perhaps better than it was."***

200 scientists responded with 'Scientists for Science', which defended GOF research and the ability to carry it out safely, but acquiesced on the possibility of further discussion, as long as it was done "under the auspices of a neutral party", such as the US National Academy of Sciences.

In a sign of the potential for common ground, Ian Lipkin, a renowned virus hunter at Columbia University in New York, saw fit to sign both calls. Both arguments have merit, he says, but both are also incomplete. Last week, a group of scientists including both opponents and supporters of GOF flu research, published a sober assessment of the potential and limitations of current approaches to assess the potential pandemic risk of various flu viruses (C. A. Russell *et al.* *eLife* 3, 03883; 2014). It paints a much more nuanced picture than some of the bold claims made earlier for GOF research. We need more such balanced analyses, and fewer dogmatic opinions, on both sides. ■

## The ice bucket

*Social-media fun for medical research bypasses animal sensitivities.*

Film-maker Steven Spielberg did it. Nobel laureates Thomas Südhof and Shinya Yamanaka did it. The fashion world's Naomi Campbell and Victoria Beckham did it. Physicist Stephen Hawking — who has the disease amyotrophic lateral sclerosis (ALS) — watched as his children did it on his behalf. They, perhaps you, and millions of others all took the 'ice-bucket challenge'.

Even if the name is unfamiliar, the images are unlikely to be. The challenge involved being filmed as you had a bucket of iced water thrown on you. For the privilege, most people pledged money for research into ALS, also known as motor-neuron disease, and then nominated others to take the challenge. The resulting little movies were posted on the Internet. It was a lot of fun.

As many of the people who took the challenge understand, ALS is a dreadful illness. Motor neurons in the brain and spinal cord degenerate

and lead to paralysis. It is relatively rare, affecting 4–7 in 100,000 people. But there is no cure, and no good understanding of its cause.

The ice-bucket challenge emerged in the United States in July and went viral around the globe, peaking in August. During that month, the ALS Association in Washington DC received more than US\$100 million in donations, compared with \$2.8 million collected during August 2013. Already, the association has distributed some \$20 million of that for research. ALS societies in Germany and the Netherlands hauled in more than \$1 million each. Australia managed more than \$2 million and Japan more than a quarter of a million. The UK Motor Neuron Disease (MND) Association in Northampton attracted 910,000 donations in just three August weeks, compared with its average monthly score of 13,000. Research has never benefited from a social-media phenomenon to this extent before.

The success of the activity is an endorsement of medical research by the general public. The associations that benefited have been careful to explain that the money will be distributed through expert review. This means that only the best research will be funded. Yet during all the excitement, what mention was made of the fact that research leading to effective treatments will eventually, one way or another, require the use of animals?

The research collaborations chosen on 2 October in the ALS Association's first round of funding are mostly based on human genomic and stem-cell approaches, which tactfully avoids the animal issue. By contrast, beneficiaries of the MND Association's windfall include both clinical research and research that uses animal models. ALS is a disease that can be caused by different factors in different people. Because its aetiology is so poorly understood, the animal models generated so far — in, for example, flies, mice and monkeys — are not totally reliable. Much will be gained from the human-genetic approaches now under way. They could help to develop better animal models.

Would members of the public have participated so joyously in the

activity if they had known that research on animals might benefit from their donations? Had that sensitive question been raised, the mood might have been different and its consequences for medical research damaged. But glossing over the reality of such research is not a good strategy for avoiding crises; instead, life scientists and their organizations should take every opportunity to say when animals have been used in research, and to explain why. Societal discussions about responsible animal research need to take place outside periods of crisis.

It is encouraging to see the tide slowly turning towards such openness — witness the MND Association's upfront funding of the full spectrum of necessary research. And outside the ice-bucket excitement, last week

***"There are many ways to support medical research."***

saw another major advance. On 13 October, the US Society for Neuroscience and the Federation of European Neuroscience Societies combined their might to publish, for the first time, a public statement in support of a neuroscientist under attack: Nikos Logothetis, a director at the Max Planck Institute for Biological Cybernetics in Tübingen, Germany, who works with monkeys. His lab had been infiltrated by an animal activist who filmed the primates there, and the videos were used as propaganda by organizations opposed to any research on animals. (An independent investigation at the institute declared that there were no systematic problems with animal care there.)

This sort of vocal support for research is important. Logothetis's work on the brain is fundamental, but applied research on degenerative diseases, including ALS, will be aided by a better understanding of the complex organ in which the diseases originate.

There are many ways to support medical research. Engaging people's enthusiasm with actions such as the ice-bucket challenge is an important one. Public support by scientific organizations for the responsible actions of their members is another. The challenge is great, the need even greater. ■

## Toxic influence

*Europe must act to stop livestock drugs from wiping out its vulture populations.*

A dead vulture in Spain could herald a crisis for raptor populations, because a drug that has killed hundreds of thousands of birds and driven some species to the brink of extinction in Asia now threatens to do the same in Europe. The European Medicines Agency (EMA) must clamp down on the drug.

The Spanish bird died two years ago. Now, the probable cause has been identified as a drug given to livestock (I. Zorrilla *et al. Conserv. Biol.* <http://doi.org/wf5>; 2014). Events in Asia show how serious the consequences could be. In the 1990s, vultures on the Indian subcontinent started dying in huge numbers. Some populations lost more than 95% of their animals. The consequences were catastrophic. As the skies cleared, dead livestock were left to rot in fields.

Research finally pinned the blame on the anti-inflammatory drug diclofenac, which had become widely used in cattle for problems ranging from pneumonia to mastitis. Although harmless to bovines, it is highly toxic to vultures that feed on the carcasses (J. L. Oaks *et al. Nature* **427**, 630–633; 2004).

As a result, India, Pakistan and Nepal placed heavy restrictions on the use of the drug in livestock. And although campaigners say that large vials officially designated for human use are often repurposed by veterinarians, the threat to the vultures of Asia has decreased. Numbers have not yet recovered, and in some cases are still declining, but the birds at least now stand a chance.

Europe is heading in the opposite direction. Despite warnings

from scientists, Spain — home to the vast majority of Europe's vultures — last year licensed diclofenac for livestock use. The EMA is considering the risks posed by the drug, and is scheduled to reach a decision by the end of November.

The discovery that the 2012 vulture was probably felled by a related drug, called flunixin (see *Nature* <http://doi.org/wfx>; 2014), is worrying for two reasons. First, it shows that diclofenac is not the only product in the class known as non-steroidal anti-inflammatory drugs (NSAIDs) that has the potential to kill vultures and other birds of prey. Second, it shows that carcasses containing significant quantities of these drugs are reaching the wild-animal food chain in Europe — in this case, probably through the Spanish tradition of wild-animal feeding stations known as *muladares*.

Two things should now happen. The EMA must move to heavily restrict — if not ban — the use of diclofenac in livestock. An alternative drug that does not harm vultures — meloxicam — is already available, and vets should use this in preference. And, as urged by the researchers who reported the flunixin-killed vulture, regulators should look at the effects of all NSAIDs used in livestock on vultures. Although diclofenac could well be the most deadly, we must know what other drugs also pose a threat to birds that feast on carrion, and how they might be managed.

In the longer term, regulators in Spain and the rest of the European Union need to ask how a drug with such evidence of environmental damage was allowed to come onto the market.

Spain is an important stronghold for vultures, and this alone would be reason enough to look seriously at restricting the use of diclofenac.

**➔ NATURE.COM**  
To comment online,  
click on Editorials at:  
[go.nature.com/xhunqv](http://go.nature.com/xhunqv)

But the European Union needs to set an example for the rest of the world. If it allows diclofenac use to continue, countries such as India could well decide to ease their restrictions, and African nations may rethink their plans to ban it. ■



TOM FINCH



## Stormy outlook for long-term ecology studies

*The closure of a 40-year project to understand and protect seabirds shows the false priorities of funders, warns Tim Birkhead.*

In the early months of this year, a series of fierce storms battered Europe's western seaboard. Seabirds struggle to feed in rough water, and some 40,000 of them soon washed up dead on beaches. Climate change is expected to increase the frequency of such storms, so to understand the impact of global warming on ecosystems, we need to analyse the long-term biological impact of these events.

Until recently, I was in an excellent position to do this. For more than 40 years, I have studied populations of guillemots on Skomer Island, off the coast of Wales. My research has revealed, for example, that the birds now breed two weeks earlier than they did in the 1970s, probably owing to climate change.

This kind of research is not easy. It has taken four decades to accumulate the data necessary to understand how the population works because to do so requires accurate measures of how long adult guillemots live, how many chicks they produce, how old they are when they breed, what proportion of young birds survive to breed and so on.

No more. Funding for the project has been axed. As it stands, I have no money to pay a research assistant to help me identify and count exactly how many of the birds have managed to survive the storms.

To assess the storms' effects, we need to gather data from the 2015 breeding season to feed into the statistical models we use to calculate survival. It is frustrating that officials chose this moment to terminate our funding, when we have such an important opportunity to assess the vulnerability of seabirds to climate change.

Guillemots are one of our most abundant seabirds, and they are excellent indicators of the quality of the marine environment. For example, they are desperately vulnerable to oil pollution, and tens of thousands have died in oil spills such as those resulting from the sinking of the *Torrey Canyon* (1967) and *Erika* (1999) oil tankers. Partly as a consequence of such disasters, guillemot numbers have fluctuated widely over the past 80 years.

In the 1930s, Skomer's guillemot population stood at around 100,000 pairs. By 1972, when I started to work with them, the numbers had fallen to just 2,000 pairs, probably owing to oil spills from ships sunk nearby during the Second World War. Since the 1980s, the numbers have increased, and there are now around 25,000 pairs.

For the past 20 years, this study — the aims of which are to understand the population biology of guillemots and to implement a scientifically robust monitoring scheme — was funded by the Countryside Council for Wales. But in 2013, the council was consumed by a new quango, Natural Resources Wales (NRW), which terminated the funding of about £12,000 (US\$19,000) per year.

NRW implied there was a shortage of cash, but I think the move was down to a change in priorities. NRW does not seem to value what my study has achieved: a comprehensive health check for guillemots. There is a feeling out there that conservation and monitoring is low-quality science and should be cheap; there is also a feeling that monitoring does not matter.

For all those biologists who start what turn out to be long-term studies, continuity of funding is a major problem. Most research grants are for 3–5 years, and in the current economic climate it is hard to predict whether funding will be renewed. Of course, all researchers dream of continuous funding, but long-term ecological studies are a special case. They are often disproportionately successful in terms of new discoveries because researchers know their system or

study species extremely well and under various environmental conditions.

Long-term population studies have shown, for example, that unlike humans, female chimpanzees do not experience a menopause. They have revealed that the age at which mute swans start and stop reproducing is a heritable trait. And they have demonstrated how rare environmental events — such as total food failure in one year — can turn cooperative, peaceful birds into selfish, brutal killers of their neighbours' offspring.

The current focus by the main funding bodies on what they consider economically useful research with a quick return is short-sighted. When my study started in the 1970s, climate change was barely on anyone's radar. The main benefit of long-term studies is that they allow

researchers to address problems that no one has yet imagined. If we are to have any hope of conserving species, we need to understand them, and we need to understand the way they are affected by environmental change.

Back in 1972, the aim of my original PhD project, supervised by Chris Perrins and the late David Lack, was to understand the dynamics of the declining population of guillemots on Skomer. Lack was famous for his work on the population biology of birds, an interest that was encapsulated in one of his best-known books, *The Natural Regulation of Animal Numbers* (1954). Quite what he thought I could achieve in a three-year PhD is still a mystery to me, given that guillemots live for at least 20 years and do not start breeding until they are at least five years old.

Forty years on, Perrins asked me whether I would soon be completing the project he set me. I would dearly like to. ■

**Tim Birkhead** is professor of zoology at the University of Sheffield, UK. e-mail: [t.r.birkhead@sheffield.ac.uk](mailto:t.r.birkhead@sheffield.ac.uk)

THERE IS A  
**FEELING**  
OUT THERE THAT  
CONSERVATION  
AND MONITORING IS  
**LOW-QUALITY**  
SCIENCE AND  
SHOULD BE CHEAP.

➔ **NATURE.COM**  
Discuss this article  
online at:  
[go.nature.com/pn9nwj](http://go.nature.com/pn9nwj)

# RESEARCH HIGHLIGHTS

Selections from the  
scientific literature

## MICROBIOLOGY

### Obesity link to jet-lagged microbes

Disrupted sleep patterns alter the composition of gut bacteria, leading to metabolic problems.

Eran Elinav at the Weizmann Institute of Science in Rehovot, Israel, and his team found that the abundance of gut microbes in mice fluctuates daily in sync with host feeding times. But when the team genetically disabled the animals' circadian clocks or shifted them by eight hours, the bacteria lost this rhythmicity and their composition changed.

Jet-lagged mice eating a high-fat diet gained more weight and showed an increased susceptibility to diabetes compared with normal mice that were fed the same food. Jet-lagged humans had more bacteria called Firmicutes — which have been linked to metabolic disease — in their guts than before their transatlantic trips.

The findings could explain why shift workers have a higher risk of obesity and diabetes.

**Cell** <http://doi.org/wfh> (2014)

## CONSERVATION

### Horn trade could save rhinos

Wild southern white rhinoceroses could go extinct in just nine years because of poaching, but could be saved if trade in their horns were to be carefully managed.

Poachers killed almost 1,000



southern white rhinoceroses (*Ceratotherium simum simum*; **pictured**) for their horns in 2013, some 5% of the total population. Enrico Di Minin of the University of Helsinki and his colleagues used population and economic models to estimate extinction risk and the cost of anti-poaching patrols.

The models suggest that the species could be saved by a carefully controlled trade in horn collected from rhinos that die naturally or harvested from live animals without killing them. Money from this would fund increased anti-poaching patrols and create an

income source for local people, deterring them from poaching. **Conserv. Biol.** <http://dx.doi.org/10.1111/cobi.12412> (2014)

## STEM CELLS

### Cell transplants enhance vision

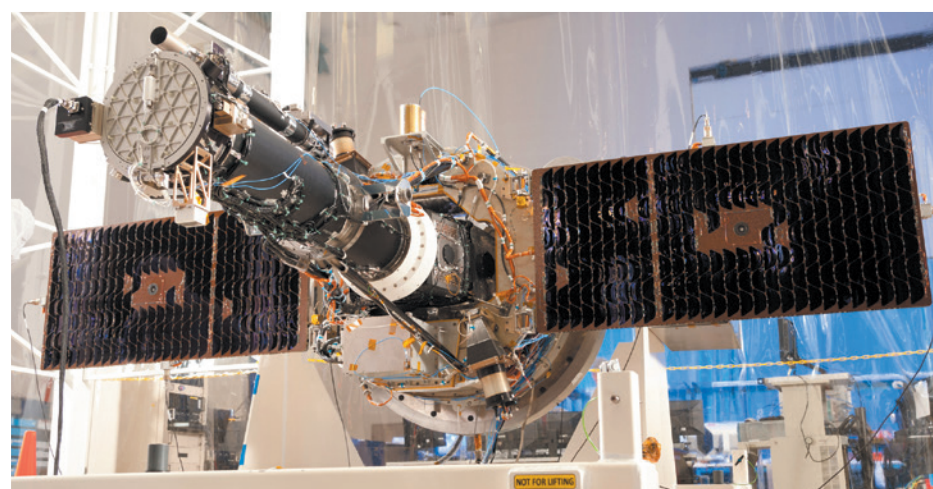
Implanted retinal cells derived from stem cells seem to be improving vision in some people in two early-stage clinical trials.

Steven Schwartz at the University of California, Los Angeles, Robert Lanza at Advanced Cell Technology in

Marlborough, Massachusetts, and their team grew retinal pigmented epithelial cells from human embryonic stem cells and transplanted them into the retinas of 18 people who have one of two forms of macular degeneration, which results in the loss of central vision.

After about two years, there have been no serious side effects from the cells, such as abnormal growth. Ten participants reported seeing more letters on an eye chart than before the treatment.

The transplanted cells are support cells that do not directly enable vision, so it is



## SOLAR PHYSICS

### Solar atmosphere is a hotbed of activity

Explosions of plasma in the Sun's atmosphere can reach temperatures of nearly 100,000°C, much hotter than scientists had expected.

The finding is one of several about the region between the solar surface and the uppermost edge of the Sun's atmosphere, or corona, revealed by NASA's Interface Region Imaging Spectrograph (IRIS) mission. The spacecraft (pictured before its launch) found that much of the energy from solar flares goes into heating and accelerating the plasma explosions, reports a team led by Hardi Peter of the Max Planck Institute for Solar System

Research in Göttingen, Germany.

Viggo Hansteen of the University of Oslo and his co-workers found short loops of magnetized plasma that flicker out within minutes and could help to explain how the corona gets so hot.

Jets of charged particles less than 300 kilometres wide also occasionally appear for up to 80 seconds, and may fuel the solar wind, say Hui Tian of the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts, and his colleagues.

**Science** <http://doi.org/wfc>; <http://doi.org/wfd>; <http://doi.org/wff> (2014)

LOCKHEED MARTIN

GARY OMBLER/DK LIMITED/CORBIS

not known how eyesight has improved. The authors could not rule out placebo and other bias effects.

**Lancet** <http://doi.org/wdf> (2014)

## PHOTONICS

## Laser moves items long distances

A laser beam can move matter tens of centimetres and in two directions.

Such tractor beams have been used to shift small objects very short distances. To scale this up, Wiesław Krolikowski at the Australian National University in Canberra and his team fired a laser beam at gold-coated hollow glass spheres in air. The photons heated up the spheres, creating a temperature difference between their far and near surfaces. This generated a force that pushed the shells in the opposite direction to the beam. By changing the beam's polarization state, the team was able to stop the spheres or reverse their direction.

The authors say that the technique could be used to gather samples remotely and for other applications.

**Nature Photonics** <http://doi.org/wft> (2014)

## METEOROLOGY

## Tornadoes growing more clustered

Tornadoes in the United States have been happening on fewer days since the 1970s, but more tornadoes have touched down (pictured) on those days.

The overall number of US tornadoes has not changed in recent decades. However, in

analysing the national tornado database, Harold Brooks of the National Severe Storms Laboratory in Norman, Oklahoma, and his team found that the number of days with at least one tornado has fallen from 150 to 100 since the early 1970s. Over the past decade, 20% of US tornadoes occurred on just three days of the year.

Whether the change is linked to rising global temperatures is not clear, the authors say.

**Science** 346, 349–352 (2014)

## NEUROSCIENCE

## Molecule boosts brain rewiring

Blocking a brain-cell receptor boosts the brain's ability to form new neuronal connections as it adapts to changing stimuli.

Carla Shatz at Stanford University in California and her colleagues disrupted the receptor, PirB, in the visual centre of mouse brains by either genetically deleting it or blocking it with a molecule.

They found that when these mice were forced to use only one eye, circuits in their visual cortices were able to rewire better than those of normal mice. This happened even in adulthood, when brain-cell rewiring becomes more difficult. In a mouse model of amblyopia, or 'lazy eye', the blocking molecule made the brain sensitive to signals from the unused eye, allowing better vision in that eye.

Targeting PirB could be a way to treat amblyopia and other brain disorders, the authors say.

**Sci. Transl. Med.** 6, 258ra140 (2014)

## PALAEOLOGY

## Strange fossil is a vertebrate cousin

Bizarre 500-million-year-old sea creatures called vetulicolians are relatives of vertebrates.

Palaeontologists have struggled to identify the relationship between living animals and these extinct

## SOCIAL SELECTION

Popular articles on social media

## Pros and cons of the PhD glut

Amid increased competition for faculty jobs in biomedicine, some have suggested cutting the number of PhD students. So when a senior scientist advised against this, the online world took notice. Eve Marder, a neuroscientist at Brandeis University in Waltham, Massachusetts, argued in the journal *eLife* that it is hard to predict who will excel in science, so any attempt to limit access to PhD programmes will inevitably exclude potential stars. The reaction was mixed. "Reduce the number of admitted graduate students? Agree with Eve Marder: not the greatest idea," tweeted Sergey Kryazhimskiy, an evolutionary biologist at Harvard University in Cambridge, Massachusetts. But Mike White, a geneticist at Washington University School of Medicine in St. Louis, Missouri, argued in a blog post that Marder was "perpetuating the PhD pyramid scheme".

**eLife** 3, e04901 (2014)



Based on data from altmetric.com. Altmetric is supported by Macmillan Science and Education, which owns Nature Publishing Group.

➔ **NATURE.COM**  
For more on popular papers:  
[go.nature.com/t9tgts](http://go.nature.com/t9tgts)

organisms, because of their odd combination of features such as gill slits and a segmented abdomen. A team led by Diego García-Bellido at the University of Adelaide and John Paterson at the University of New England in Armidale, both in Australia, analysed a fossil vetulicolian from a South Australian island.

The fossil, a new species named *Nesonektris aldridgei*, shows the outline of a notochord — a rod-like structure that develops into the backbone in vertebrates.

Although *N. aldridgei* is distantly related to vertebrates, its closest relatives are tunicates — invertebrates that swim or attach themselves to underwater rocks. It was probably a free-swimming filter-feeder, say the authors.  
**BMC Evol. Biol.** 14, 214 (2014)

## CANCER

## Immunotherapy beats leukaemia

Engineering certain immune cells to kill cancerous cells in leukaemia has driven the disease into remission for up

to two years in more than half of participants in an early-stage clinical trial.

Stephan Grupp at the Children's Hospital of Philadelphia, Pennsylvania, and his co-workers tested their approach on 30 people with acute lymphoblastic leukaemia, including 25 children, who had failed to respond to conventional treatment or relapsed.

The team engineered a patient's T cells to express a receptor that targets the cancerous B cell, and infused the T cells back into the person. After one month, 27 people were in remission, and after up to 2 years, 78% survived — a much higher rate than with chemotherapy. Those in remission had high blood levels of the engineered T cells.

However, all of the participants had inflammatory side effects that required hospitalization.

**N. Engl. J. Med.** 371, 1507–1517 (2014)

➔ **NATURE.COM**  
For the latest research published by Nature visit:  
[www.nature.com/latestresearch](http://www.nature.com/latestresearch)



MIKE HOLLINGSHEAD/CORBIS



# SEVEN DAYS

The news in brief

## RESEARCH

### Beyond Pluto

Planetary researchers have identified three possible targets for NASA's New Horizons spacecraft after it flies past Pluto next year. Mission scientists used the Hubble Space Telescope to search the Kuiper belt in the distant reaches of the Solar System for objects that the probe could reach with its limited amount of fuel. Of the three, the best candidate, dubbed PT1, is several tens of kilometres across. If NASA approves funding for the visit, New Horizons would fly by PT1 in late 2018 or early 2019.

## EVENTS

### Quake trial appeal

On 18 October, an Italian court heard opening arguments from defence attorneys appealing against the conviction of six scientists and a government official found guilty of manslaughter over the earthquake that hit L'Aquila, Italy, on 6 April 2009. On 31 March 2009, the scientists had been part of an expert panel on seismic risk that prosecutors say resulted in a reassuring message being released to the public. According to the charges, that message led to the death of L'Aquila residents who stayed in their homes when they would otherwise have left. See [go.nature.com/xf4smu](http://go.nature.com/xf4smu) for more.

### Ebola update

A US\$1-billion appeal by the United Nations to help fight Ebola in West Africa has yielded only about \$376 million in pledges, the organization said on 16 October. Last week, the World Health Organization (WHO) estimated that mortality rates had neared 70%, and warned that Liberia,

Sierra Leone and Guinea could see up to 10,000 new cases per week by December. The WHO also declared Nigeria and Senegal free of Ebola after 42 days had passed since their last reported case. In the United States, President Barack Obama on 17 October appointed former White House adviser Ron Klain to oversee the country's response to Ebola.

## PEOPLE

### Plea offer

Prosecutors have offered a plea deal to Dong Pyou Han that could resolve his fraud case without a trial. The former Iowa State University scientist is accused of misconduct in

an AIDS-vaccine study. In court documents filed on 16 October, Han's defence team asked for more time to consider the terms, which must be translated into Korean, and requested postponement of the trial, set for 3 November. In December 2013, the US Office of Research Integrity reported finding that Han spiked rabbit blood samples with antibodies to produce false results that were widely disseminated.

### Costly case

The University of California, Los Angeles, spent almost US\$4.5 million on defending chemist Patrick Harran against criminal charges stemming from a research assistant's

death in a 2008 fire in his lab, according to documents revealed by the *Los Angeles Times* on 16 October. The publicly funded university, which has since spent more than \$20 million on improving lab safety, says that taking on the legal expenses was within its rights and obligations. In June, Harran agreed to a deal that could result in charges being dismissed (see [go.nature.com/vpdwvz](http://go.nature.com/vpdwvz)).

## POLICY

### Pathogen pause

The US White House Office of Science and Technology Policy on 17 October announced a mandatory moratorium



C. ONEZIA/SIF

## Snail find ruffles feathers

The rediscovery of a colourful snail (pictured), declared extinct in 2007, on Aldabra atoll in the Seychelles Islands has revived tensions over the study published in *Biology Letters* that reported the animal's disappearance (J. Gerlach *Biol. Lett.* 3, 581–585; 2007). At the time, biologist Clive Hamblin of the University of Oxford, UK, and his colleagues penned a comment that criticized the study's methods, contested the claims of extinction and requested a retraction. But the comment was rejected for publication by the

journal, and the Intergovernmental Panel on Climate Change has since cited the snail as a prime example of species loss caused by climate change. Conservation workers, however, rediscovered the Aldabra banded snails (*Rhachistia aldabrae*) in a remote part of the atoll in August 2014. On 15 October, editor-in-chief Richard Battarbee said in an editorial that the journal had invited Hamblin to resubmit his comment, but that the invitation had been declined.

JOE CAVARETTA/AP

on new funding for gain-of-function research, which increases the deadliness of pathogens such as the virus that causes influenza. The moratorium suspends government funding for such work, and asks scientists who have already been funded to pause their research voluntarily while two non-regulatory bodies — the National Science Advisory Board for Biosecurity (NSABB) and the National Research Council — assess the risks. See pages 403 and 411 for more.

## Australian edge

The Australian government unveiled a national strategy on 14 October for strengthening ties between industry and science, and increasing commercial returns on public research funding. The 'competitiveness agenda' established the Commonwealth Science Council, a group of ten appointed experts from business and research, which will become the leading body for advising the government on science and technology. Five centres will be set up, at a projected cost of Aus\$188.5 million (US\$165 million), to improve public-private science collaborations, and to boost Australia's mining, oil and medical technologies industries. See [go.nature.com/ydbkdd](http://go.nature.com/ydbkdd) for more.



## Nuclear waste site

In a long-delayed safety evaluation, the US Nuclear Regulatory Commission said on 16 October that designs for a proposed national nuclear-waste disposal site at Yucca Mountain in Nevada (**pictured**) contained the necessary barriers for long-term isolation of radioactive waste. Congress designated the site as a potential repository in 1987, and in 2008 the Department of Energy applied for a construction licence. But the project has been stalled for years by political opposition and legal challenges (see *Nature* 473, 266–267; 2011). Last week's report spurred renewed calls by Republicans to revive the project.

### BUSINESS

## Fusion dreams

Defence contractor Lockheed Martin announced ambitious plans for nuclear fusion on

15 October, with a ten-year road map to commercialize a reactor small enough to fit on a lorry. Lockheed's Skunk Works division in Palmdale, California, has so far been tight-lipped about the results of its years of research, and several outside experts have responded to the company's lofty goals with scepticism. See [go.nature.com/u597zk](http://go.nature.com/u597zk) for more.

## Buyout backdown

The biopharmaceutical company AbbVie seems to be backing away from a US\$54-billion bid to take over Shire, a drug manufacturer based in Dublin. On 15 October, AbbVie's board of directors recommended that stockholders in the firm, based in North Chicago, Illinois, reject the deal in the event of a vote. The company cited recent changes to US tax rules that eliminated some of the anticipated financial benefits of the transaction. If the deal does not go through by 14 December, AbbVie will pay Shire a \$1.6-billion break fee.

## Genomics boost

Genetic-analysis company Illumina announced on 15 October the selection of three start-up firms to help to develop biotechnology applications for next-generation sequencing. Encoded Genomics,

## COMING UP

### 27–31 October

The Intergovernmental Panel on Climate Change meets in Copenhagen to adopt a report that synthesizes findings of its Fifth Assessment Report, and to approve the report's summary for policy-makers.

### 29–31 October

In Aix-en-Provence, France, the 3rd International Conference on Biodiversity and the UN Millennium Development Goals tackles sustainability and food security. [go.nature.com/e8snpu](http://go.nature.com/e8snpu)

EpiBiome and Xcell Biosciences are the first participants in the Illumina Accelerator programme, which will give each company a US\$100,000 loan; lab and office space in San Francisco, California; and extra support for experiments and reagents. See [go.nature.com/tynw7o](http://go.nature.com/tynw7o) for more.

## Desertec dries up

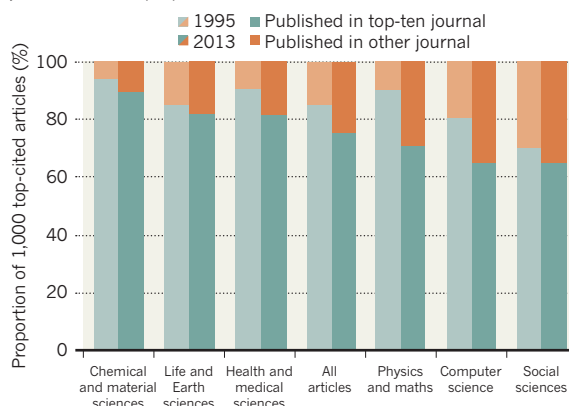
Following the withdrawal of most of its shareholders, the Desertec Industrial Initiative (Dii) has buried ambitious plans to supply Europe with power from solar plants and other renewable sources across North Africa and the Middle East. Only three of its existing 19 shareholders remained after a meeting last week in Rome. Saudi Arabia's ACWA Power, Germany's RWE and the State Grid Corporation of China plan to remodel Dii into a service company for facilitating regional renewable-energy projects in North Africa and the Middle East.

## TREND WATCH

An analysis by Google suggests that, compared with 1995, a greater proportion of the 1,000 top-cited papers now appear in publications other than the ten top-cited journals (A. Acharya *et al.* Preprint at <http://arxiv.org/abs/1410.2217>; 2014). Phil Davis, a publishing consultant in Ithaca, New York, says that the trend could be partly due to an ever-increasing number of articles being published coupled with a decline in the number of articles being published in elite journals (see [go.nature.com/qbztbe](http://go.nature.com/qbztbe)).

## ELITE JOURNALS LOSING DOMINANCE?

Most of a field's top-cited articles are still published in its top-cited journals, but the proportion has declined since 1995.



SOURCE: GOOGLE/ARXIV

**NATURE.COM**

For daily news updates see:  
[www.nature.com/news](http://www.nature.com/news)

# NEWS IN FOCUS

**POLITICS** US election campaign promises dim funding prospects **p.412**

**ANTHROPOLOGY** Oldest modern-human genome casts light on African exodus **p.413**

**ASTRONOMY** Planet-hunting mission long over, Kepler keeps going **p.414**



**ARCHAEOLOGY** Remembering the day that *Homo floresiensis* was found **p.422**

CHINA FOTOPRESS/GETTY



Outbreaks of influenza have prompted research into strains with pandemic potential.

## BIOSECURITY

# US suspends risky disease research

*Government to cease funding gain-of-function studies that make viruses more dangerous, pending a safety assessment.*

BY SARA REARDON

**T**he US government surprised many researchers on 17 October when it announced that it will temporarily stop funding new research that makes certain viruses more deadly or transmissible. The White House Office of Science and Technology Policy is also asking researchers who conduct such 'gain-of-function' experiments on influenza, severe acute

respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS) to stop their work until a risk assessment is completed — leaving many unsure of how to proceed.

"I think it's really excellent news," says Marc Lipsitch, an epidemiologist at the Harvard School of Public Health in Boston, Massachusetts, who has long called for more oversight for gain-of-function research. "I think it's common sense to deliberate before you act."

Critics of such work argue that it is unnecessarily dangerous and risks accidentally releasing viruses with pandemic potential — such as an engineered H5N1 influenza virus that easily spreads between ferrets breathing the same air<sup>1,2</sup>. In 2012, such concerns prompted a global group of flu researchers to halt gain-of-function experiments for a year (see *Nature* <http://doi.org/wgx; 2012>). The debate reignited in July, after a series of lab accidents involving mishandled pathogens at the US Centers for Disease Control and Prevention in Atlanta, Georgia.

The White House's abrupt move seems to be a response to renewed lobbying by gain-of-function critics who wanted such work suspended and others who sought to evaluate its risks and benefits without disrupting existing research.

Arturo Casadevall, a microbiologist at the Albert Einstein College of Medicine in New York City, calls the plan "a knee-jerk reaction." "There is really no evidence that these experiments are in fact such high risk," he says. "A lot of them are being done by very respectable labs, with lots of precautions in place."

Some researchers are confused by the moratorium's wording. Viruses are always mutating, and Casadevall says that it is difficult to determine how much mutation deliberately created by scientists might be "reasonably anticipated" to make a virus more dangerous — the point at which the White House states research must stop. The government says that this point will be determined for individual grants in discussions between funding officers and researchers.

One of the most prominent laboratories conducting gain-of-function studies is run by Yoshihiro Kawaoka, a flu researcher at the University of Wisconsin–Madison. In 2012, Kawaoka published a controversial paper<sup>1</sup> reporting airborne transmission of engineered H5N1 flu between ferrets. He has since created an H1N1 flu virus using genes similar to those from the 1918 pandemic strain<sup>3</sup>, to show how such a dangerous flu could emerge. The engineered H1N1 was transmissible in mammals and much more harmful than the natural strain.

Kawaoka says that he plans to comply with the White House directive to halt current research once he understands which of his projects it affects. "I hope that the issues can be discussed openly and constructively so that important research will not be delayed indefinitely," he says.

But it seems that the freeze could be lengthy. The White House says that it will wait for ►



► recommendations from the US National Science Advisory Board for Biosecurity (NSABB) and the National Research Council before deciding whether and how to lift the ban. The groups are expected to finish their work within a year. As *Nature* went to press, the NSABB was set to convene on 22 October, its first meeting in two years. Lipsitch, who will speak at the event, says that he will advocate for the development of an objective risk-assessment tool to evaluate individual research projects. In particular, he says, decision-makers should consider whether a gain-of-function study makes a contribution to a public-health goal, such as the prevention and treatment of flu, that could

justify both the risk and the use of money that could be spent on safer research.

"There clearly are going to be instances where gain-of-function research is necessary and appropriate, and there are others where the opposite applies," says Ian Lipkin, a virologist at Columbia University in New York City. The need to understand the ongoing Ebola outbreak in West Africa and control its spread, for instance, emphasizes the importance of infectious-disease research — as well as the regulation of such work, Lipkin says. Although public worry about Ebola being transferred through the air is unfounded, researchers could make a case for the need to determine how the virus

could evolve in nature by engineering a more dangerous version in the lab. "I think we should have some sort of guidelines in place before such experiments are even proposed," says Lipkin. Yet Ebola is not included in the White House's research-funding ban, and a spokesperson says that there are no plans to include it on the list.

One thing is certain, says Casadevall: the NSABB meeting is certain to see heated debate as scientists from all sides convene. "I hope they have enough room," he adds. ■ [SEE EDITORIAL P.403](#)

1. Imai, M. *et al. Nature* **486**, 420–428 (2012).
2. Herfst, S. *et al. Science* **336**, 1534–1541 (2012).
3. Watanabe, T. *et al. Cell Host Microbe* **15**, 692–705 (2014).

## POLITICS

# US midterm elections offer little hope for science

*November vote is unlikely to break a political stalemate that has squeezed research funding.*

BY LAUREN MORELLO

When US voters head to the polls on 4 November, they are poised to set in motion a major political shift that promises to intensify partisan strife over issues such as climate change, immigration and research funding. For the first time since 2006, Republicans are likely to win full control of the US Congress — having seized the House of Representatives in 2010, they are now predicted to take control of the Senate.

The development seems inauspicious for US researchers who depend on government funding. Prominent Republicans have repeatedly questioned the veracity of biological evolution and human-induced climate change, and party leaders' push for drastic spending cuts has resulted in across-the-board reductions known as sequestration, which slashed 5.1% from science agencies' budgets in 2013.

Yet in fact, the changing balance of power is expected to have little practical impact — because Congress may not be able to do much of anything. Experts see little hope of breaking the political gridlock that has made the current Congress, which began in January 2013, arguably the least productive in modern history. "It doesn't matter what happens to the Senate," says Michael Lubell, director of public affairs for the American Physical Society in Washington DC. "The outcome is going to be the same."

Although Republicans already hold a commanding advantage in the House of Representatives, they are expected to win only a simple



The deadlocked US Congress has been one of the least productive in modern history.

majority in the 100-member Senate, not the three-fifths majority that generally enables a party to pursue its legislative agenda without drawing any minority-party support.

That is a double-edged sword. Congress is unlikely to approve large increases in funding. But Republicans will have a hard time pushing through bills to enact the more extreme parts of their agenda — such as blocking new federal regulations to cut carbon emissions, or a plan to require the National Science Foundation to certify that all of its grants serve the 'national interest'. (That proposal, from the House science

committee chairman Lamar Smith (Republican, Texas) is targeted mostly at funding for research in the social and behavioural sciences.) And even if such legislation were approved by the House and Senate, President Barack Obama would almost certainly exercise his veto power.

Relations are so bad between the two major parties that even the US pledge to help stop the Ebola outbreak in West Africa has sparked bickering. Earlier this month, Senator James Inhofe (Republican, Oklahoma) and several high-ranking Republicans in the House temporarily blocked Obama's request for

DAVID BURNETT/CONTACT PRESS IMAGES

## PALAEOANTHROPOLOGY

# Oldest-known human genome sequenced

*DNA shows a group of modern humans roamed across Asia.*

BY EWEN CALLAWAY

A 45,000-year-old leg bone from Siberia has yielded the oldest genome sequence for *Homo sapiens* on record — revealing a mysterious population that may once have spanned northern Asia. The DNA sequence from a male hunter-gatherer also offers tantalizing clues about modern humans' journey from Africa to Europe, Asia and beyond, as well as their sexual encounters with Neanderthals.

His kind might have remained unknown were it not for Nikolai Peristov, a Russian artist who carves jewellery from ancient mammoth tusks. In 2008, Peristov was looking for ivory along Siberia's Irtysh River when he noticed a bone jutting from the riverbank. He dug it out and showed it to a police forensic scientist, who identified it as probably human.

The bone turned out to be a human left femur, and eventually made it to the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany, where researchers carbon-dated it. "It was quite fossilized, and the hope was that it might turn out old. We hit the jackpot," says Bence Viola, a palaeoanthropologist who co-led the study of the remains. "It was older than any other modern human yet dated." The luck continued when Viola's colleagues found that the bone contained well-preserved DNA, and they sequenced its genome to the same accuracy as that achieved for contemporary human genomes (Q. Fu *et al. Nature* 514, 445–449; 2014).

The researchers named their find Ust'-Ishim, after the district where Peristov found the remains. They dated him to between 43,000 and 47,000 years old, nearly twice the age of the next-oldest known complete modern-human genome, although older, archaic-human genomes exist.

DNA may be the only chance to connect the remains to other humans. "This guy came out of nowhere — there's no archaeology site we could connect it to," says Viola, suggesting that his group roamed far and wide.

The Ust'-Ishim man was probably descended from an extinct group that is closely related to humans who left Africa more than 50,000 years ago to populate the rest of the

world, but later went extinct, Viola says.

The most intriguing clue about his origin is that about 2% of his genome comes from Neanderthals. This is roughly the same level that lurks in the genomes of all of today's non-Africans, owing to ancient trysts between their ancestors and Neanderthals. The Ust'-Ishim man probably got his Neanderthal DNA from these same matings, which, past studies suggest, happened after the common ancestor of Europeans and Asians left Africa and encountered Neanderthals in the Middle East.

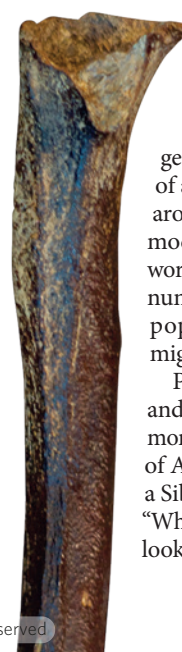
Until now, the timing of this interbreeding was uncertain — dated to between 37,000 and 86,000 years ago. But Neanderthal DNA in the Ust'-Ishim genome pinpoints it to between 50,000 and 60,000 years ago on the basis of the long Neanderthal DNA segments in the Ust'-Ishim man's genome. Paternal and maternal chromosomes are shuffled together in each generation, so that over time the DNA segments from any individual become shorter.

The more precise dates for Neanderthal-human mating pose a challenge for scientists who have proposed that modern humans left Africa before 100,000 years ago and reached Asia more than 75,000 years ago, says Chris Stringer, a palaeoanthropologist at London's Natural History Museum. Those researchers, who include Michael Petraglia, an archaeologist at the University of Oxford, UK, have pointed to *H. sapiens*-like bones from the Levant that are older than 100,000 years and to 70,000-year-old

stone tools found in India as evidence for an early human exodus to Asia along a southern coastal route that eventually reached Oceania and Australia.

But Petraglia sees Ust'-Ishim's genome differently. "I think this is part of a population boom that's going on around 45,000 years ago, which means modern humans got to the ends of the world by 45,000 years ago," he says. Their numbers might have swamped human populations that arrived in earlier migrations.

Petraglia expects that ancient DNA and other fossil finds will paint a much more complicated picture of the peopling of Asia. "This is just a random find in a Siberian river deposit," Stringer says. "What else could be there when they start looking systematically?" ■



The Ust'-Ishim femur.

US\$750 million to send military personnel and other resources to the outbreak zone.

Funding for many science agencies did rise slightly overall in the 2014 fiscal year after the mandatory sequestration in 2013. The National Oceanic and Atmospheric Administration's budget actually rose by roughly \$575 million in 2014, to \$5.3 billion. But the National Institutes of Health (NIH), the world's largest biomedical-research funder, received \$29.9 billion, less than its pre-sequester budget of \$30.7 billion.

"Everybody in Congress knows that science is important," says Congressman Rush Holt (Democrat, New Jersey), "but they don't have much appreciation of what it takes to sustain it." (Physicist Holt, one of the few scientists in US public office, famously gives out bumper stickers that read, "My Congressman IS a Rocket Scientist!")

The NIH cuts have put the squeeze on many research institutions. At the Washington University School of Medicine in St Louis, Missouri, for example, federal grant funding fell by 14% from fiscal year 2012 to 2013, and by another 3% from 2013 to 2014. "Part of it was sequestration, part of it was the timing of some big grants expiring," says Jennifer Lodge, the university's vice-chancellor for research. "It sort of reeks of an atmosphere of gloom around funding."

The story is similar at the University of Arizona in Tucson, which relies heavily on support from the NIH and NASA. Although the university has begun seeking more funding from industry and philanthropists, "there's still a sense of unease about what the future looks like," says Kimberly Andrews Espy, the university's senior vice-president for research.

Others are more optimistic. The University of Maryland in College Park saw its share of federal grants rebound in the 2013–14 academic year, rising nearly 3% on the year before. Its chief research officer Patrick O'Shea says that success in chasing large awards has helped his institution to prosper as federal budgets have grown tighter. "Rather than complaining, we're trying to be more effective," he says.

Beyond the coming election, US policy analysts are already looking to another, key political test: whether lawmakers will finalize a 2015 budget deal before they depart for the year in late November. Jennifer Zeitzer, deputy director of public affairs at the Federation of American Societies for Experimental Biology in Bethesda, Maryland, sees signs of a "good-faith effort" to approve a spending deal in the next several weeks.

But even if the lawmakers deliver, she says, that will not make up for four years of legislative torpor. "This Congress was so spectacularly unproductive," she says, "that even showing up will give the next Congress a leg up." ■





The Rho Ophiuchi cloud, a star-forming nebula, is being observed by the Kepler space telescope.

## ASTRONOMY

# Sun's stroke keeps Kepler online

*Space telescope beats mechanical failures to begin a second mission that will trace new celestial targets.*

BY MARK ZASTROW

The crippled Kepler space telescope is unexpectedly enjoying a second lease of life. The exoplanet-hunting probe will now cast its gaze on star clusters, the centre of the Milky Way and the Solar System's outer planets as it scans a ribbon of the cosmos for the next three years. This month it has been gazing at gas clouds shrouding infant stars in the constellations Scorpius and Ophiuchus.

The telescope, originally designed to look for Earth-like planets orbiting other stars in our

region of the Galaxy, has just yielded the first data set since its reincarnation after mechanical failures. Its science team is still busy analysing data from the initial planet-hunting mission, so Kepler's managers at NASA have left it to the wider astronomical community to choose specific targets for a second mission, known as K2, and to comb through the output.

For a patched-up instrument, it is doing remarkably well. Scientists are just getting to grips with the first K2 observations, made available last month — and they say that the data are promising, leaving them eager for more.

When Kepler launched in 2009, it sought to answer one question: how common are other Earths in the Galaxy? The telescope stared at roughly 150,000 stars near the constellations Cygnus and Lyra, monitoring their brightness and watching for a momentary dimming that would signify planets crossing in front of them. Four years yielded more than 4,000 potential planets, hundreds of them Earth-sized and rocky. At least one of those bodies was found in its star's habitable zone, where liquid water could exist on its surface (see *Nature* <http://doi.org/wf4>; 2013).

Crucial to Kepler's success were its four reaction wheels, which held it steady like a gyroscope as it circled the Sun. It survived the loss of one wheel in July 2012. But when a second failed in May 2013, Kepler was down to two wheels, with only two axes of control.

"Sorrow, disappointment, a little grief in there," is how John Troeltzsch, Kepler programme manager at Ball Aerospace, says he felt when he found out about the second failure. His team had built and operated the craft for NASA at Ball's facilities in Boulder, Colorado. Kepler's mission seemed to be over. Then, three days later, he opened an e-mail from Ball engineer Doug Wiemer, proposing a fix. Five hours and eight e-mails later, Wiemer had outlined a plan to get Kepler back on its feet.

Wiemer had fashioned a crutch for Kepler using the only resource available: sunlight. Positioned so that its long side faces the Sun, the spacecraft leans against the pressure created by the onslaught of photons and balances using its two good wheels. With this approach, the team hoped to get within a factor of ten of Kepler's original performance — but with additional software refinements, NASA's Kepler project manager Charlie Sobeck says that it is better than that, more like a factor of two or three. Wiemer thinks that further tweaks will close the gap entirely.

One limitation of the K2 mission is that Kepler must keep the Sun side-on as it orbits, forcing the telescope to switch its field of view roughly every 80 days. This is not enough time to hunt for Earth-like planets around Sun-like stars, but it does let K2 track other celestial bodies such as clusters of newly-formed stars (see 'Changing focus').

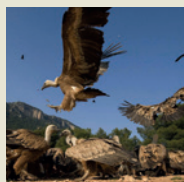
In February, Kepler will turn its gaze to the well-known clusters Pleiades and Hyades, followed in April by the Beehive and M67 clusters.

NASA/JPL-CALTECH/HARVARD-SMITHSONIAN CFA



**MORE  
ONLINE**

## TOP NEWS



Cattle drug awakens fears for safety of Spanish vultures  
[go.nature.com/2rx3av](http://go.nature.com/2rx3av)

## MORE NEWS

- Mars orbiters survive close brush with comet [go.nature.com/9bprc6](http://go.nature.com/9bprc6)
- Hints of 'axion' dark matter from the Sun [go.nature.com/qwsj2t](http://go.nature.com/qwsj2t)
- Chinese scientists charged with misuse of 'megaproject' money [go.nature.com/pic6dg](http://go.nature.com/pic6dg)

## NATURE PODCAST



Exocomets, exotic oxidations and extraordinary *Deinoceratops* dinosaurs [nature.com/nature/podcast](http://nature.com/nature/podcast)

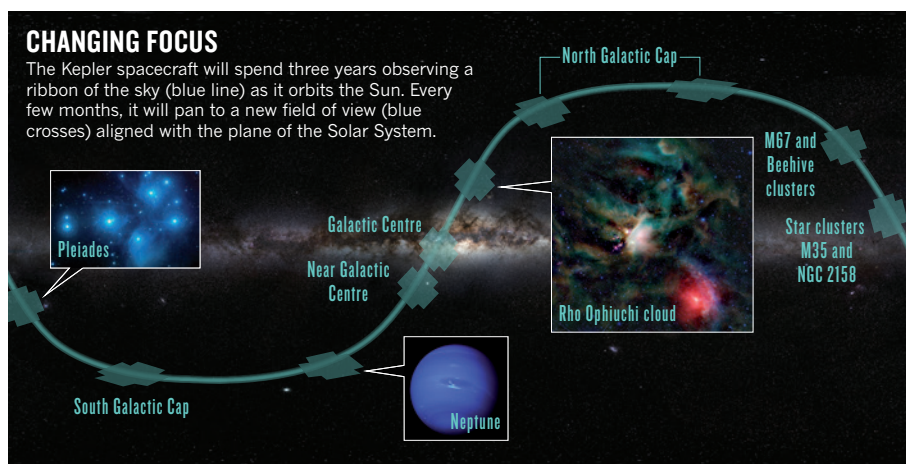


NASA

Astronomers love these objects because the stars within them are the same age, which is easily deduced by plotting their brightness and colour. The observations should offer snapshots of planetary systems during their early development, which could resolve debates about how planets form and migrate.

Other opportunities lie closer to home. In Kepler's previous life, it discovered that Neptune-sized ice giants are the most common planets in the Galaxy. This year and next, it will point at the Solar System ice giants Neptune and Uranus, hoping to learn more about their internal structures by observing the flickers of seismic vibrations.

The team has planned as far ahead as April 2016, when Kepler will look towards the centre of the Milky Way in search of mysterious bodies called free-floating planets. Past observations suggest that Jupiter-sized planets outnumber stars in the Galaxy by a factor of two or more. Most can be detected only when they cross in front of a distant star and the planet's gravity bends the star's light like a lens. Kepler should be able to confirm the population of these loners, says Andrew Gould, an astrophysicist at Ohio State University in Columbus. Presumably, some free-floating planets were kicked out of their systems, but they are so massive that it is hard to imagine how, he says; one



possibility is encounters with other stars.

Other potential targets abound, including brown dwarfs — the smallest stars, which have clouds and storms like Jupiter — and white dwarfs, which are dim compared with Sun-like stars. This makes any planets hosted by white dwarfs easier to see, and thus they are tempting targets for future observatories such as NASA's James Webb Space Telescope, which is scheduled to launch in 2018.

Some astronomers worry that K2's productivity will be limited not by the availability of data, but by NASA's research grants. Only

US\$2 million per year is guaranteed for Kepler work; a separate \$17-million pot is spread annually across all analyses of data archived from previous NASA missions. The move to increase archival funding is a slow culture shift for NASA, says Gregory Sloan, an astronomer at Cornell University in Ithaca, New York, but one that is necessary to give researchers the years they will need to continue reaping scientific benefits from missions such as K2. "The thing is, it's a huge, rich data set," he says. "And there will be years and years before we have really digested it." ■

# Data bank struggles as protein imaging ups its game

*Hybrid methods to solve structures of molecular machines create a storage headache.*

BY EWEN CALLAWAY

Structural biology, the mapping of complex biological molecules such as proteins, is in the grip of a revolution. The field has long been dominated by X-ray crystallography, a technique made iconic by its role in decoding the DNA double helix in the 1950s. But the need to tackle more complex structures and to watch ‘molecular machines’ function in real time is fuelling a shift towards hybrid imaging methods that can create moving models.

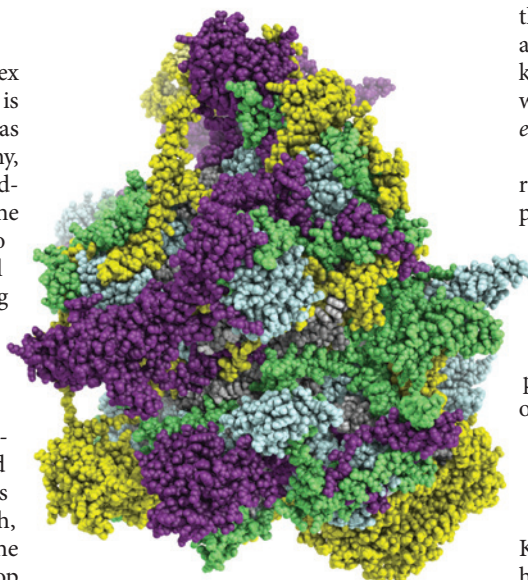
That is posing a challenge for the world’s official repository for protein structures: the Protein Data Bank (PDB), which relies almost exclusively on crystallography data and lacks the standards and software infrastructure to archive structures described by hybrid methods. This month, leaders of the four organizations around the world that host the data bank held a workshop in Hinxton, UK, to hatch a plan to ensure that hybrid models and their insights into fundamental biology and disease do not get lost.

Historically, structural biology has focused on generating three-dimensional (3D) descriptions of individual proteins. In many cases, this is a task perfect for crystallography, in which a molecule is bombarded with X-rays and the pattern of scattered radiation reveals the position of each atom. The technique underpins dozens of discoveries that led to Nobel prizes.

Archiving structures in the centralized, free PDB is crucial because it enables other researchers to use them to address questions never imagined by their discoverers. Most journals will publish structures only if they have been deposited in the PDB. This year, the database topped 100,000 registered structures, the vast majority of which were determined using X-ray crystallography (see *Nature* **509**, 260; 2014).

But in the past decade or so, structural biology has moved on. Researchers now want to describe intricate cellular structures made up of dozens, or even hundreds, of proteins that move relative to each other do jobs such as recycling proteins or copying chromosomes. These molecular machines cannot be coaxed into the tidy, immobile crystals required for X-ray

**“These days, being a crystallographer is not good enough.”**



**A subunit of a ribosome, a molecular machine.**

crystallography. “These days, being a crystallographer is not good enough,” says Gerard Kleywegt, a structural biologist at the European Bioinformatics Institute in Hinxton, who heads the European annex of the PDB.

Hybrid methods take an ‘everything but the kitchen sink’ approach to structural biology, incorporating many different techniques. Some can offer a dynamic view of a molecular machine in motion; for example, fluorescence resonance energy transfer measures the distance and interactions between proteins. Others, such as cryo-electron microscopy, can deliver near-atomic detail of entire complexes without the need to crystallize them. Computer programs then integrate the various bits of information — including data from crystallography-friendly proteins inside the molecular machine — to produce a 3D model that best fits the data.

The scientific literature is now studded with products of the hybrid approach. In 2012, structural computational biologist Andrej Sali of the University of California, San Francisco, and his collaborators used hybrid methods to describe the structure of the 26S proteasome complex (K. Lasker *et al. Proc. Natl Acad. Sci. USA* **109**, 1380–1387; 2012), which recycles proteins and may malfunction in neurodegenerative diseases such as Alzheimer’s. The researchers have now used the model to identify potential drugs

that alter the proteasome’s activity. This year, another team published a hybrid model of the key HIV proteins that sneak the virus into a cell, which may help in vaccine design (M. Pancera *et al. Nature* <http://doi.org/wfz>; 2014).

The hybrid approach has also tackled the ribosome, which produces proteins; the nuclear pore complex, which provides a gateway between the genome in the nucleus and the rest of the cell; and the molecular syringes made by bacteria that inject proteins into cells. Models of many more molecular machines are expected. “We’re going to enter a period of exponential growth in the generation of these hybrid structures,” says Stephen Burley, a structural biologist at Rutgers University in Piscataway, New Jersey, who heads one of the two US annexes of the PDB.

At the PDB workshop, on 6–7 October, Kleywegt, Burley and three dozen others hashed out the challenges that these triumphs are creating for the PDB. Crystallography yields a standardized set of data files in which a structure and its level of precision are self-evident; by contrast, the underlying data for the hybrid models exist in a mishmash of formats such as X-ray diffraction patterns or electron-micrograph pictures. And going from raw data to a model involves more steps with hybrid methods than in crystallography; it also requires more assumptions, often leading to multiple possible ways of interpreting the results.

Most workshop attendees agreed that it will be crucial for structural-biology databases to capture not just the hybrid models’ raw data, but also how the models were put together, so that other researchers can verify and build on them. But there are many questions, such as how to store and distribute the data sets, which are much larger than crystallography files. The meeting ended with an agreement to seek funding for a new bank centred on molecular machines — and to come up with a name for it.

It is imperative to find a way to curate hybrid structures if structural biology is to realize its potential, says cell biologist Jan Ellenberg of the European Molecular Biology Laboratory in Heidelberg, Germany, who led one of the teams that modelled the nuclear pore complex. Ultimately, he says, “we want to have the molecular structure of an entire cell. That’s still science fiction at the moment — but it’s somewhere we can get to in 10, 20 years.” ■

B. J. GREBER ET AL. NATURE HTTP://DOI.ORG/WG4 (2014).

# THE ETHICS SQUAD

*Bioethicists are setting up consultancies for research — but some scientists question whether they are needed.*

BY ELIE DOLGIN

**S**tacy Hodgkinson and Amy Lewin had the best of intentions when they enrolled the pregnant 15-year-old in their study. The psychologists were evaluating an educational programme for young parents-to-be, and the teenager met all the inclusion criteria: she was 15–32 weeks pregnant with her first child, under 19 years of age, and her partner — who did not live with her — was willing to participate in the study. There was just one problem. Dad was 24 years old, and according to local laws he was guilty of child sexual abuse for sleeping with a minor.

The couple had apparently lied to each other about their ages, but not to Hodgkinson and Lewin, both then at the Children's National Health System in Washington DC. This presented a dilemma. The scientists had promised the participants that their information would be kept confidential. But did that trump their legal duty to report the crime to the police? And how would that affect the family?

"Here was a young father telling us he'd like to be involved in his child's life in a positive way," says Lewin, who is now at the University of Maryland in College Park. Telling the authorities, she says, "could potentially do more harm than good".

In search of moral and legal guidance, Hodgkinson and Lewin contacted Tomas Silber, a paediatrician who also runs a research ethics consultation service, a 'one-stop shop' for advice on thorny research issues.

To Silber, the course of action was clear.

"There's only one thing you can do," he says. "You have to report it." After explaining their legal obligations to the couple, Lewin and Hodgkinson told the police, who launched an investigation. The teen and her partner broke off contact with the researchers, and Hodgkinson does not know whether the father maintained a positive presence in the child's or the mother's life — which was ultimately the goal of their programme. "Sometimes you do the right thing, but the consequences aren't good," says Silber.

Ethical dilemmas in research are nothing new; what is new is that scientists can go to formal ethics consultancies such as Silber's to get advice. Unlike the standard way that scientists receive ethical guidance, through institutional review boards (IRBs), these services offer non-binding counsel. And because they do not form part of the regulatory process, they can weigh in on a wider range of issues — from mundane matters of informed consent and study protocol to controversial topics such as the use of experimental Ebola treatments — and offer more creative solutions.

The consulting services are "a really new area", says Joshua Crites, a research ethicist at the Pennsylvania State College of Medicine in Hershey. "Even some of the most basic questions get complicated really quickly, and it's better to have a group of ethicists working together to sort this out."

But many scientists either do not know that they exist or fear using them because they could add red tape to an already heavy

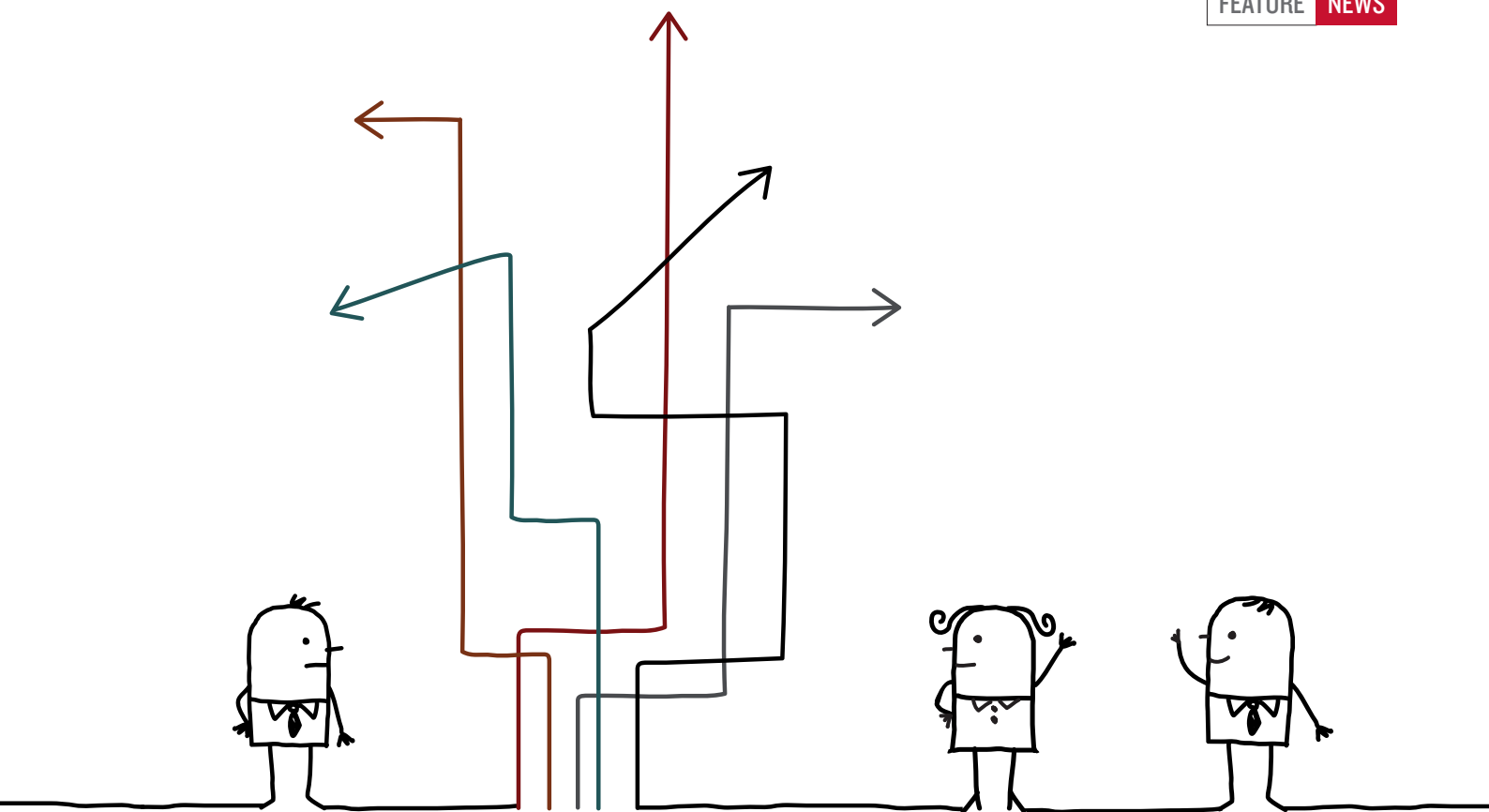
administrative burden. And this year, the US National Institutes of Health (NIH) scrapped funding for a working group to support ethics-consultation services and to develop best practices for the profession.

Although financial support could return in some form, ethicists are not waiting around for it. Benjamin Wilfond, director of the Treuman Katz Center for Pediatric Bioethics at Seattle Children's Hospital in Washington, has set up the Clinical Research Ethics Consultation Collaborative, a group of around 35 bioethicists who hope to keep improving the consultation service model, even without NIH support. "There's energy behind continuing what we started," says Holly Taylor, a research ethicist at the Johns Hopkins Berman Institute of Bioethics in Baltimore, Maryland, and a member of the group.

## HERE TO HELP

IRB approval is required for almost all human-subject research in the United States. The foundations for current IRB practices emerged 40 years ago in the wake of numerous ethical lapses in research, including the infamous Tuskegee experiments performed in Alabama between 1932 and 1972, in which doctors allowed syphilis to progress untreated in hundreds of African American men. Today, IRBs are the main channels for policing ethics in academic medical studies. But their primary function is to ensure adherence to regulatory and legal requirements. They do not always include members with bioethics expertise, and





NLSHOP/SHUTTERSTOCK

discussion of ethics sometimes takes the form of box-ticking rather than careful deliberation.

That is where consultants come in. Unlike IRBs, consultants can provide guidance throughout a study — not just at the point of regulatory review — and do so in a non-confrontational advice-giving capacity. They offer “an open space for talking about research ethics in a way that is not driven by the regulatory environment”, says Marion Danis, chief of the bioethics consultation service at the NIH Clinical Center, a research hospital in Bethesda, Maryland.

The Clinical Center was the first organization to launch a research ethics consultancy, in 1996, and a handful of academic medical centres followed suit over the next decade. Then, in 2006, the NIH launched the Clinical and Translational Science Award programme to enhance drug development and testing in academic settings, and it led to a rapid expansion of the concept in the United States. According to a survey published last year, by 2010 more than 30 academic institutions had set up research-ethics consultation services. That said, fewer than half of them had fielded calls by researchers seeking advice in the previous year, and just six got more than ten calls<sup>1</sup>. “In most places, these have not ended up being high-volume activities,” says Steven Joffe, a medical ethicist who led a fairly idle service at Harvard Medical School in Boston, Massachusetts, until moving to the University of Pennsylvania Perelman School of Medicine in Philadelphia in 2013.

**“You need some independent person to say, ‘Well, let’s step back and think about this.’”**

that began there in 2003. Her team needed to obtain informed consent from participants, but the researchers on the ground in Iraq were concerned that including the University of Washington’s name on the consent forms — a requirement for IRB approval — would make it difficult to get the data they needed. “They feared that being associated with American institutions would get them killed”, says Hagopian. “They dug in their heels and refused” to carry the form.

Hagopian wanted to strip the university’s name from the consent document, but the IRB insisted that it was an important part of informed consent, which is meant to protect participants, not the investigators. The impasse brought Hagopian and her team to Wilfond. He concluded that it would be ethical to remove mention of the institution, for three main reasons: first, research subjects would also be placed at risk by signing a document linking them to the University of Washington; second, apart from the link to the United States, the

Amy Hagopian, a global-health researcher at the University of Washington in Seattle, found herself turning to an ethics consultant for help with a study in Iraq to find out how many people had died as a result of the US-led conflict

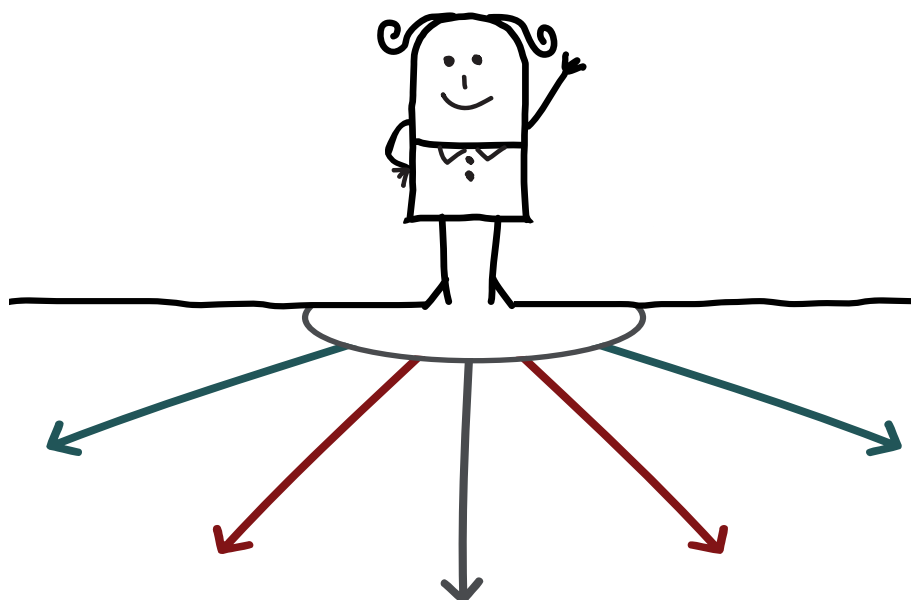
research involved minimal risk to the participants; and third, the study would not happen unless the name of the institution was removed.

The IRB eventually agreed with Wilfond. The researchers went ahead with the study and found that nearly half a million people had died from causes attributable to the Iraq war between 2003 and 2011 — a figure much greater than most previous estimates<sup>2</sup>. “We couldn’t have done this without him,” Hagopian says of Wilfond.

#### WORLDLY ADVICE

Of course, bioethicists have been providing advice about research for years, long before the NIH created a formal service. Outside the United States, ethics consultations mostly happen through the regional equivalent of an IRB or take place in casual conversations or ‘kerbside consults’. “All in all, it’s pretty ad hoc,” says Mark Sheehan, who studies ethics at the Ethox Centre of the University of Oxford, UK.

At some institutions in Canada, ethics advice about research studies can also be sought through the services that help patients and doctors to settle end-of-life decisions and other moral issues in health care. Unlike in the United States, where training programmes in research ethics and clinical ethics are usually separate, in Canada “we all tend to have both kinds of expertise pretty much”, says Ann Heesters, a bioethicist at the Toronto Rehabilitation Institute in Ontario, one of the only Canadian hospitals that publicizes the availability of ethics consultations for researchers. According



to Heesters, around one in every seven of her consultations pertains to research.

In Australia, “it’s very difficult for researchers to be able to seek advice before they submit the full application” for official ethics review, says Nikola Stepanov, who studies research ethics and law at the University of Queensland in Brisbane. And if a human-research-ethics committee — the Australian equivalent of an IRB — finds ethical problems in a study’s protocol, researchers may have trouble finding a formal channel for further guidance.

“We’re obviously in the stage that the United States was at before it brought in these ethics consultations,” says Stepanov. “Something more formalized would be very appropriate.”

But not all ethicists agree that a separate service is needed, even within the United States. “If the IRB has the responsibility for ethics review, why are we pulling in someone else?” asks Susan Kornetsky, director of clinical research compliance at Boston Children’s Hospital in Massachusetts. Norman Fost, who studies ethical and legal issues in research at the University of Wisconsin–Madison, would rather see bioethics panels folded into the standard IRB structure. Because IRBs are “a toll gate that everybody has to go through,” Fost says, these panels, which would ideally include qualified ethicists, should “look at every single protocol and identify problems that nobody else has yet identified”. Relying on a separate, optional service means that some problems could be missed. “It’s the cases they’re not getting called about that worry me,” he says.

### COMPLEMENTARY SERVICES

Advocates say that the aim of consultancy services is to complement IRBs and other oversight bodies, not to become entwined with them. “For innovative research designs, you need some independent person to say, ‘Well, let’s step back and think about this not just

**“It’s the cases they’re not getting called about that worry me.”**

from the standpoint of do the regulations permit it, but does it fulfil the spirit of what people want done with the public research enterprise?” says bioethicist Steven Miles at the University of Minnesota in Minneapolis. Wilfond has been working to increase the visibility and the rigour of ethics consultancies. Last year, for example, he and Taylor launched a biannual series in the *American Journal of Bioethics* entitled ‘Challenging Cases in Research Ethics’. The latest case, from Silber and his colleagues describing the obligation to report statutory rape, was published in September<sup>3</sup>. Wilfond is also collecting descriptive data about consultations and has expanded the reach of his service at the University of Washington by welcoming external requests — including from pharmaceutical companies, which typically employ armies of lawyers but rarely bioethicists. In such cases, the University of Washington consults on a fee-for-service basis: US\$200 an hour for drug companies, less for non-profit organizations.

The Stanford Center for Biomedical Ethics in California also works with drug firms. There, panellists provide their time and advice at no cost, on the condition that they can publish case studies. In 2011, for example, a start-up company approached the centre for guidance on the sale and promotion of a prenatal genetic test that involves analysing fetal DNA circulating in maternal blood (see *Nature* 478, 440; 2011). The consultation led to an academic paper that called for amendments to informed-consent procedures and restrictions on the sale of direct-to-consumer tests<sup>4</sup>.

“Many of our consults end up that way,” says

Mildred Cho, associate director of the Stanford centre. “We do treat these things as scholarly activity as well as a service.” Cho estimates that around one-quarter of her service’s cases come from the drug industry.

Wilfond is currently working to expand the panels to draw in a wider range of views and to broaden the experience of panellists, a move that he considers one of his most innovative for ethics consultancies. In June, he was called into a meeting at Seattle Children’s Hospital with Ron Gibson, director of the hospital’s cystic fibrosis centre. Gibson had been gathering data from several studies that were using laboratory tests that can be performed only in a research setting or fall outside of recommended guidelines, but he was unsure whether he should incorporate the results into patients’ routine clinical care. Seven bioethicists from Wilfond’s collaborative telephoned into the meeting, ready to offer their take.

As the consultation began, Wilfond explained that the point of bringing the ethicists into the discussion was twofold. First, it would offer Gibson a wider range of opinions, and second, it would expose the advisers on the phone to a case they might not otherwise have been involved in. “There’s a lot of learning that goes on bidirectionally,” Wilfond says. The hour-long meeting was “educational”, says Gibson, who has since implemented a new policy for his research programme, although he declined to discuss specifics. “The spectrum of opinions on various levels of data sharing was reassuring that there is likely not one best way to address the issue.”

Wilfond and his colleagues hope that more scientists and clinicians will start to see the benefits of their services. “There just hasn’t been an awareness of how important this is,” says Charles MacKay, a consultant in clinical and research bioethics in Bethesda, Maryland.

But getting scientists to actually buy into such services may require a shift in attitudes. “Researchers generally have become members of a culture of research compliance,” says Christian Simon, a bioethicist at the University of Iowa Carver College of Medicine in Iowa City. They are responsive to what IRBs require, he says, but that sometimes means that they are unwilling to step back and consider the finer ethical details.

“We’re not the ethics police,” says Reid Cushman, co-director of the ethics consultation service at the University of Miami in Florida. “We’re just another resource to help you stay out of trouble.” ■

**Elie Dolgin** is a science writer in Somerville, Massachusetts.

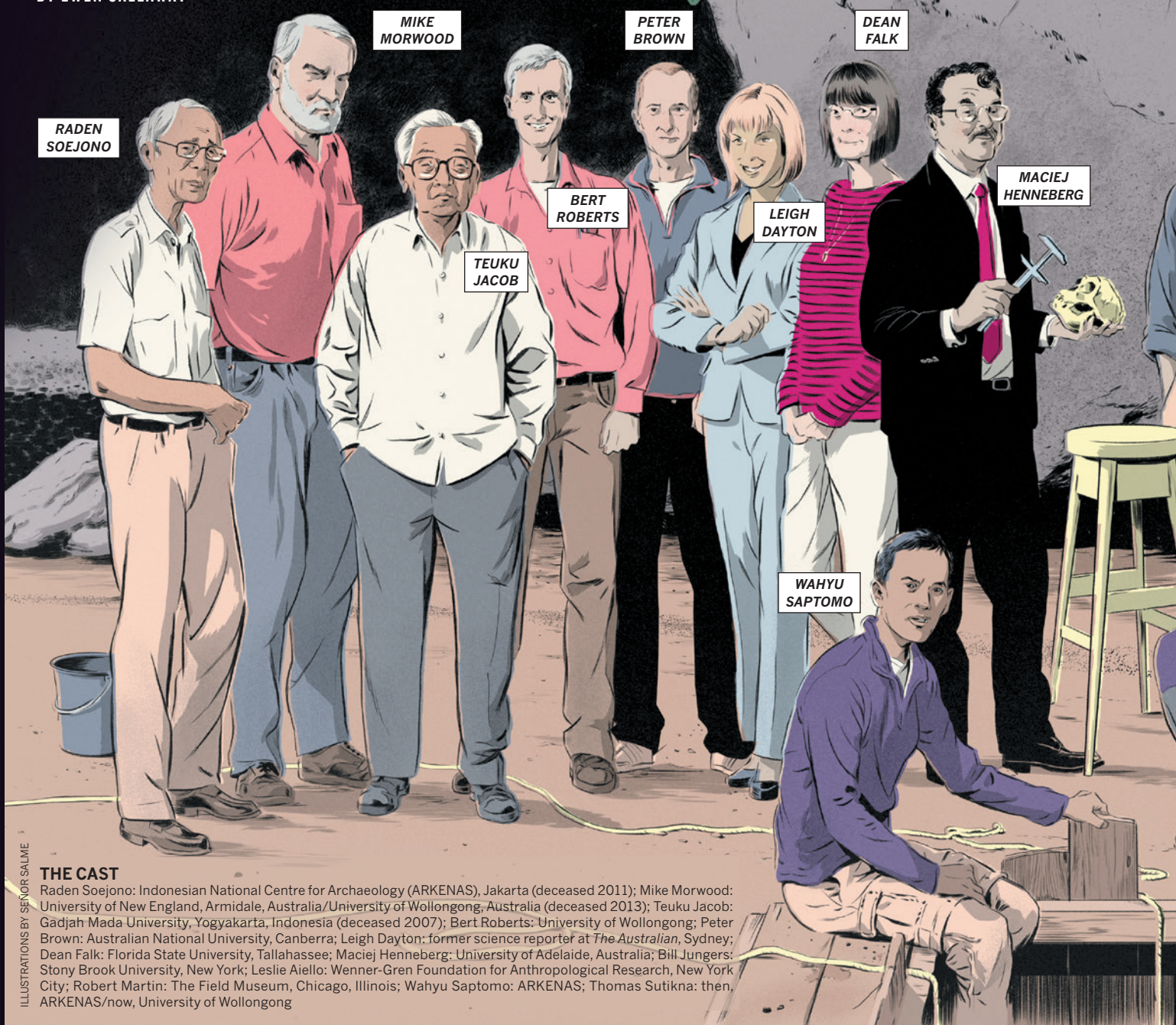
1. McCormick, J. B. *et al. Clin. Transl. Sci.* **6**, 40–44 (2013).
2. Hagopian, A. *et al. PLoS Med.* **10**, e1001533 (2013).
3. Hodgkinson, S., Lewin, A., Chang, B., Beers, L. & Silber, T. *Am. J. Bioeth.* **14**, 54–55 (2014).
4. Allyse, M. A. *et al. Prenat. Diagn.* **33**, 656–661 (2013).



# Tales of the hobbit

In 2004, researchers announced the discovery of *Homo floresiensis*, a small relative of modern humans that lived as recently as 18,000 years ago. The 'hobbit' is now considered the most important hominin fossil in a generation. Here, the scientists behind the find tell its story

BY EWEN CALLAWAY



## THE CAST

Raden Soejono: Indonesian National Centre for Archaeology (ARKENAS), Jakarta (deceased 2011); Mike Morwood: University of New England, Armidale, Australia/University of Wollongong, Australia (deceased 2013); Teuku Jacob: Gadjah Mada University, Yogyakarta, Indonesia (deceased 2007); Bert Roberts: University of Wollongong; Peter Brown: Australian National University, Canberra; Leigh Dayton: former science reporter at *The Australian*, Sydney; Dean Falk: Florida State University, Tallahassee; Maciej Henneberg: University of Adelaide, Australia; Bill Jungers: Stony Brook University, New York; Leslie Aiello: Wenner-Gren Foundation for Anthropological Research, New York City; Robert Martin: The Field Museum, Chicago, Illinois; Wahyu Saptomo: ARKENAS; Thomas Sutikna: then, ARKENAS/now, University of Wollongong





**T**he hobbit team did not set out to find a new species. Instead, the researchers were trying to trace how ancient people travelled from mainland Asia to Australia. At least that was the idea when they began digging in Liang Bua, a large, cool cave in the highlands of Flores in Indonesia. The team was led by archaeologists Mike Morwood and Raden Soejono, who are now deceased.

**THOMAS SUTIKNA (field archaeologist in charge of the excavation):** In 1999, Mike came to our office and proposed excavating at Liang Bua. 'Liang Bua' means cold cave. It's 500 metres above sea level, and it's situated very close to the confluence of two rivers, which provide natural resources like water and raw materials for stone artefacts. The roof is really high, providing good circulation. There's regular sunlight year round. It's very suitable for habitation.

**RICHARD 'BERT' ROBERTS (geochronologist who conceived the dig with Morwood):** The excavations started off on a very small scale in 2001, but we found some interesting things: bones of stegodons, which are these now-extinct primitive elephants. There were lots of Komodo dragons, lots of rat bones, all sorts of other species, including this kind of giant stork. We didn't find anything spectacular until 2003.

**WAHYU SAPTOMO (field archaeologist):** Before Mike Morwood left for the season in 2003, I said, "Why are you leaving now? If you leave, maybe we will find something important." A few days later, on 2 September, I was supervising sector VII. Our local workers were digging at around 5.9 metres. Their trowel met with a skull. A member of our team who specializes in animal and human bones came down and said, "Yes, I'm sure that's a human bone. But it's very small." Thomas, he was sick and was at the hotel that day. So I went back and met with him. I said, "We have something very important. We found the first hominid in the Pleistocene layer."

**SUTIKNA:** Immediately, my fever vanished. I couldn't sleep well that night. I couldn't wait for sunrise. In the early morning we went to the site, and when we arrived in the cave, I didn't say a thing because both my mind and heart couldn't handle this incredible moment. I just went down to the pit and looked at the bones carefully. It would be impossible to get them out because of the condition of the bones. So we decided to cut the remains out, together with the sediment, block by block, and bring them back to the hotel. We needed several days to take out all the bones.

**ROBERTS:** It was a very small body. That was the first thing that was immediately apparent — but also an incredibly small skull. We first thought, "Oh, it's a child." There was a guy who was working with us called Rokus. He did all the faunal identifications of the bones. But Rokus said, "No, no, no, it's not a child. It's not modern human at all. It's a different species."

**SAPTOMO:** Thomas drew the skeleton on paper, and he faxed the drawing to Mike and to Professor Soejono in Jakarta.

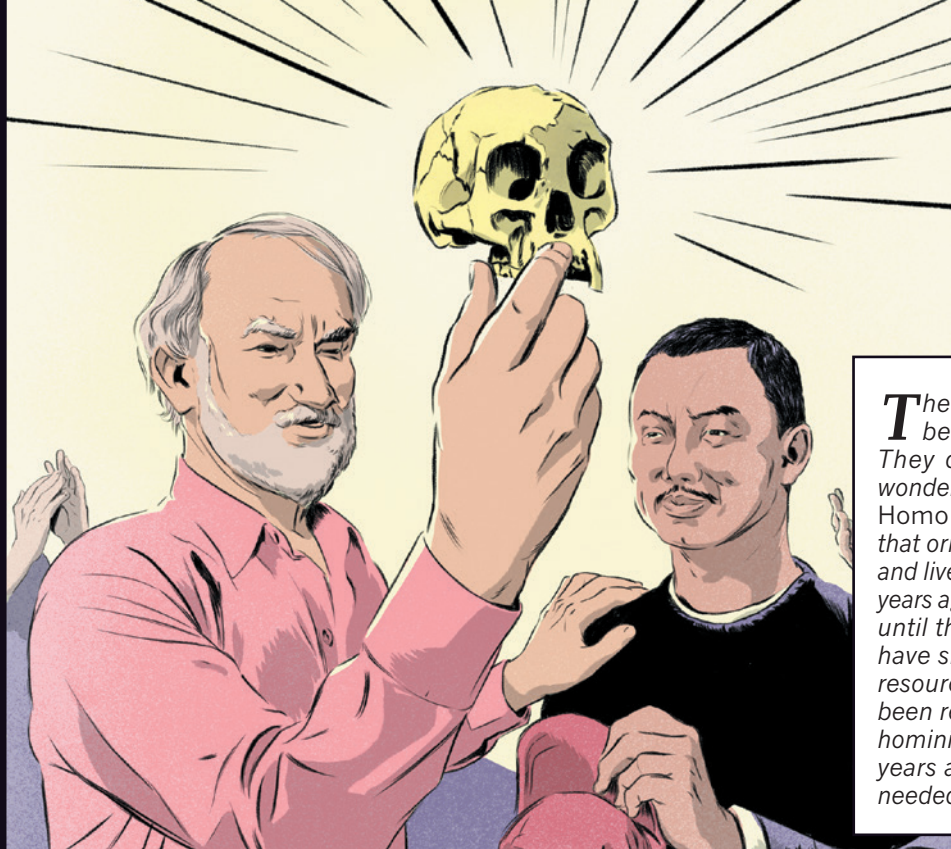
**SUTIKNA:** Mike called me at night. I couldn't understand what he was saying over the phone, he was so excited.

**ROBERTS:** Mike invited Peter Brown to come and look at the remains. Peter's a very good palaeoanthropologist, but he's kind of a difficult person as well. Peter can be kind of prickly.

**PETER BROWN (palaeoanthropologist):** Mike doesn't know much about human skeletons, and the Indonesian researchers didn't either. I was quite sceptical. The drawing may as well have been a Greek urn in terms of looking like anything much at all.

I was interested and willing to go to Jakarta. It's an interesting place to visit. I like the food. I like the atmosphere and the culture and everything else, but I didn't expect to find anything interesting or important. At the most, I thought it was going to be a sub-adult modern human ►





primitive-looking human who was living this side of the last glacial maximum, this side of the last Ice Age.

**BROWN:** If Mike had said he'd found evidence of an alien spaceship on Flores, I would have been less surprised.

*The team soon determined that the skeleton belonged to a female just over a metre tall. They dubbed her LB1. Brown and Morwood wondered whether the species was an offshoot of Homo erectus — an ancient relative of humans that originated in Africa about 2 million years ago and lived on Java, near Flores, until about 150,000 years ago. If its descendants had survived on Flores until the tail end of the last Ice Age, they might have shrunk in response to the island's limited resources. Alternatively, the species might have been related to australopithecines, small-bodied hominins that roamed Africa more than 2 million years ago. Brown and Morwood knew that they needed to let the world in on their find.*

► skeleton, probably dating to the Neolithic period or maybe a little bit earlier. The other possibility was a pathological individual, someone with a growth disorder. Those were my expectations when I turned up.

**ROBERTS:** Peter's as sceptical as I was, probably thinking, "A new human species? Sure, probably Mike getting overexcited in Jakarta. He'd been in the bush for too long." Good on him for flying over there straight away, because most people have got teaching commitments and things you've got to be getting on with.

**BROWN:** I walked into the laboratory with Mike and the lower jaw — the mandible — had been cleaned. And it was in about six seconds, maybe less, of looking at the lower jaw, I knew it couldn't have been a modern human lower jaw. I knew it had to be from another species, and things went on from there. I started cleaning the skull and doing other work on the collections. Everything was very, very soft and had to be dried out and coated with preservative. It would have been very easy to scratch or smash. If you'd stepped on it, you would have ended up with a pile of mashed potato, more or less.

**ROBERTS:** Some people, like the guys in Africa, seem to work on things for about 10 or 15 years before you finally get a fossil description. Peter was working at lightning speed by comparison. To Mike and myself it still seemed to take forever.

**BROWN:** I smuggled some mustard seeds through customs for the purpose of measuring the volume of the brain. So I cleaned it all as carefully as I could. I turned it upside down, and I poured the seeds in it. I'd taken enough seeds to measure the size of a modern human brain, say 1.5 litres of seeds, but it only took about 400 millilitres. I was flabbergasted. The last time things with a brain that size walked was around about 2.5 million to 3 million years ago. It was not making any sense at all. I recorded it a second time, a third time. Mike and Thomas are looking at me and wondering why I'm going a bit pale. I was trying to push more seeds into the skull with my finger to try and increase the volume, because it was insane really.

**ROBERTS:** The carbon dates came in and they were around 18,000 years. So at that point it was, "Oh, this is absolutely bizarre." This was a very

**HENRY GEE (senior editor at Nature):** I had no warning. Usually with these things you tend to get a little bit of scuttlebutt. But this one just came onto my desk one day in March 2004, and there it was.

**ROBERTS:** Poor old Henry probably fell off his chair when he got the papers.

**GEE:** I have to say at first it didn't strike me as this most fantastic discovery. They had this strange creature and the tone of the paper was very subdued. When you're an editor you read between the lines, and the line was: "Help us. We don't know what this thing is. We're just going to describe it and we're going to give it a very non-committal name and see what you think."

**BROWN:** I thought it was a new species and probably a new genus. I just thought it was so different.

**GEE:** When it came to us, they had given it this Latin name, *Sundanthropus floresianus* — man from the Sunda region from Flores. Well, the referees said it's a member of *Homo* so that's what it should be, and one of the referees says *floresianus* actually means 'flowery anus' so it should be *floresiensis*. So *Homo floresiensis* came along.

**ROBERTS:** We knew we had to come up with a name for publicity purposes. We couldn't call it *Homo floresiensis*, so Mike said, "I like hobbit." I said, "Okay as long as it's not going to cause any problems with Tolkien's estate," or whatever they're called. They can get pretty stropy with people using their trademarked words. Mike referred to LB1 as hobbit, not 'the' hobbit, as if its name was Mary. For a while, Mike was trying to persuade Peter Brown to call it *Homo hobbitus*. I think he just thought Mike was a complete charlatan for even suggesting it.

**BROWN:** Mike and I didn't agree about nicknames because I thought it trivialized it, and I thought it would result in every loon on the planet telephoning me as soon as it was published. And that was true — endless bizarre telephone calls from people who had seen some small hairy person in their backyard.

**MORE ONLINE**

See Nature's online special on the hobbit: [go.nature.com/hobbit10](http://go.nature.com/hobbit10)

**W**hen the papers reporting the discovery were published<sup>1,2</sup> on 27 October 2004 (28 October in Australia and Indonesia), they grabbed public attention in a way few science stories manage.

**LEIGH DAYTON (science correspondent):** It was huge, absolutely huge. Everybody was talking about it. Even my editors who absolutely do not like science whatsoever, they were fascinated. I'm just looking at the newspaper, the hard copy of the story I wrote for *The Australian*, and the other stories are all the usual political stuff, police probes, inflation figures, and then, "Small, but they're only human".

**BILL JUNGERS (palaeoanthropologist):** I had to check the date to make sure it wasn't April Fool's Day. It was so preposterous on the surface that there could be this little hominin that evolved in isolation in southeast Asia for God knows how long and persisted until almost the Holocene.

**ROBERTS:** This one really did garner just a massive amount of media interest. Way over the top. Every newspaper wanted to talk to you, TV programmes; everyone wanted to talk to everyone.

**BROWN:** The press being the way they are, they always like controversy. It's no good just to have a good story. Nobody wants to read that, so they're always trying to find someone who disagrees.

**MACIEJ HENNEBERG (palaeoanthropologist):** I had a phone call at 7 a.m. on 28 October 2004, from an Australian Broadcasting Corporation (ABC) journalist, who asked me, "What do you think about the new find?" I said, "I don't think anything, you just woke me up." He said it was published in *Nature* that there is this new species. I said, "Okay. Give me a few hours so I can find the papers." While I was reading, I was reminded of a paper on a microcephalic [small-brained] skull from Crete, about 4,000 years old. All the measurements of the LB1 skull were not significantly different from this clearly pathological skull from Crete. So at 11 a.m., I went on ABC radio and I said I think that what was found was a pathological specimen. This kind of explanation attracted a lot of attention.

**F**urther controversy erupted when Teuku Jacob, head of Indonesia's national palaeoanthropology institute, decided that the hobbit's bones belonged in his lab.

**ROBERTS:** Soejono invited Teuku Jacob to come over and have a look at the bones, and then Jacob just put them in a suitcase and walked out the door with them. Mike was absolutely ballistic. I didn't think we'd ever actually get to see the bones back again.

**BROWN:** The real disreputable thing was they tried to take moulds and cast this material. I hadn't done that because it was clear the material was too soft and too fragile to take moulds. When they had done that, the lower jaw was broken, the skull was damaged.

**T**he bones made it back to Jakarta, but the debate over the hobbit's identity grew hotter. Morwood brought in specialists to examine the fossil, and they agreed with him that it was a new species. Some key studies focused on endocasts — moulds of the inside of the hobbit's skull that revealed details of its brain.

**JUNGERS:** Mike didn't have to ask twice. I was introduced to the team in person in Jakarta in 2006, and a good part of my career has been obsessing about this fossil ever since.

**ROBERTS:** Quite a few on the American side started to throw their weight behind our team, and that helped steady the ship. They really took the hobbit to pieces and put it back together again, and found it was really a very unusual kind of animal.

**JUNGERS:** I was able to assemble a nearly complete foot that was unlike anything I've ever seen in the fossil record. I think these little guys were climbers. I don't know if you've ever been to Flores, but there were these huge Komodo dragons on the island when those guys were around. The adults don't climb, so if I were a hobbit, I would find refuge in the trees.

**DEAN FALK (evolutionary anthropologist):** Mike Morwood invited me to prepare and describe the endocast. I had a bias going into the study. I thought, because the brain was so small, that it was going to look like other primates with brains of that size, namely apes, and it didn't. It didn't look like a chimp brain. What it mostly looked like in overall shape was the endocast from *Homo erectus*.

**O**ther scientists have continued to support the idea that the LB1 specimen was a diseased human.

**ROBERT MARTIN (biological anthropologist):** I think there is something seriously weird about the LB1 specimen. The best I could come up with is microcephaly. There are hundreds of genes that can produce a small brain, with knock-on effects throughout the body.

**FALK:** We felt like, okay, we need to do a microcephalic study. I ended up running with that ball and got together a very small sample, about ten, but it's really hard to find ten endocasts from microcephalics. ▶





► We looked at LB1 and we showed there was no way it was a microcephalic. As far as I'm concerned, that paper<sup>3</sup> answered it, and I think that persuaded most everybody, except a few people. I think they even were persuaded eventually, because they changed diseases. The 'sick-hobbit hypothesis', Bill Jungers called it.

**JUNGERS:** It seems like we've got a new one every day. It's just crazy, crazy stuff. We spent a lot of time unfortunately having to deal with things like Laron syndrome and cretinism and other wild and woolly hypotheses.

**HENNEBERG:** About two and a half years ago, it all clicked. I could see that all signs of the bones were compatible with Down's syndrome. There are about 20 or so characteristics that are matching. There is not a single characteristic of LB1 that doesn't match.

**MARTIN:** I don't think we've made much progress in ten years, quite honestly. What we have is entrenched positions. We should be talking about the interpretations and the facts, not casting aspersions. I'm not an idiot because I'm questioning this.

**LESLIE AIELLO (palaeoanthropologist):** There were some important issues raised during the controversy that the proponents of a new-species hypothesis had to address. But the criticism of the hypothesis didn't turn out to hold any water.

**FALK:** In palaeoanthropology there's always controversy over new specimens. Always has been, always will be. I was a little surprised that in this day and age it would be so over the top.

**AIELLO:** There were personalities involved. The field is full of a lot of egos, and particularly male egos.

**ROBERTS:** I think the view now is: yup, it's not a diseased modern human. But whether it's a shrunk down version of a *Homo erectus*, or whether it's something more ancient like a *Homo habilis*, or even an australopithecine who's managed to struggle out of Africa — that's still pretty much up for grabs.

**BROWN:** I'm mostly interested in how it got to be where it was, which will require the discovery of additional material. That may not happen in my lifetime.

**AIELLO:** It'll happen. I tell my students that something is found every year and I never gave the same lectures twice.

**ROBERTS:** That's why we're still trying to dig up in the centre of the island, looking in the Soa Basin on Flores. Mike took another view: let's try and find the bones of the ancestors, wherever they came from, which was probably north of Flores. So Mike and I went to the Philippines, and we also went to Sulawesi, Indonesia. Mike was still doing excavations on Sulawesi, as are several of the people here.

**JUNGERS:** I never met anybody who was as single-minded and sustained in their work as Mike. He was always looking over the horizon to the next excavation, the next expedition. I excavated at Liang Bua off and on, and the last time I saw Mike was that summer before he passed.

**ROBERTS:** Mike actually came and saw me and said, "Ah Bert, I need to talk to you about something." He said "I've got cancer." He seemed to be getting tired more and more easily, and that wasn't like Mike at all.

**JUNGERS:** Mike died of complications of prostate cancer, and he was so consumed by his work that I think he neglected his health. It didn't occur to him that he should take better care of himself. Even when he was diagnosed, the only thing he wanted to talk about was the next expedition.



He was an original and became a good friend, and I miss him.

**ROBERTS:** Who would have thought ten years ago that Mike wouldn't have been with us now? He was one of the forces of nature. Without Mike, it wouldn't have happened.

*The hobbit team is still digging up Liang Bua. With new dating work, the researchers hope to determine when *H. floresiensis* went extinct and whether it overlapped with modern humans in the region. The hobbit's discovery thrust southeast Asia to the forefront of research into human evolution, suggesting that key events might have happened there. But the find also complicated the history of Homo species in Asia.*

**ROBERTS:** We had such a nice simple story, where we had modern humans and Neanderthals, and we bumped them off, that was the end of Neanderthals. We ventured across southeast Asia and it was basically empty because *Homo erectus* had died out there already, and we sort of just wandered into Australia and there we go. It was a clean and almost crisp little story. It made nice sense. Everyone was happy with that. And then suddenly the hobbit pops its head up.

**BROWN:** Now I'm more open to the idea that very small-bodied and small-brained bipeds moved out of Africa at a much earlier date, maybe 3 million years ago, or earlier. I'm more open to the idea that there were lots of failures in the evolution of bipeds. Some were successful, some weren't. It's a very branchy tree, and it just so happens we've survived.

**ROBERTS:** To me, the ultimate value of the hobbit is not what it is, in and of itself, because it's just a dead end. It probably didn't lead to anything that's alive today. But it opened up the door for people to think more broadly about everything. I think the hobbit changed the way people thought. ■ SEE COMMENT P.427

Ewen Callaway writes for Nature from London.

1. Brown, P. et al. *Nature* **431**, 1055–1061 (2004).
2. Morwood, M. J. et al. *Nature* **431**, 1087–1091 (2004).
3. Falk, D. et al. *Proc. Natl Acad. Sci. USA* **104**, 2513–2518 (2007).



# COMMENT

**EMERGENCIES** Hurricane Sandy experience led to disaster training **p.430**



**HISTORY** Mary Sommerville, first popular science author, reappraised **p.432**

**CLIMATE CHANGE** Defenders of 2°C global-warming goal respond **p.434**

**EARTH OBSERVATION** China gifts free access land-cover map to United Nations **p.434**

JIM WATSON/AFP/GETTY



The adult skull of *Homo floresiensis* (centre) at the 2004 press conference announcing the species' discovery.

## Small remains still pose big problems

Ten years after the publication of a remarkable find, **Chris Stringer** explains why the discovery of *Homo floresiensis* is still so challenging.

In early 2004, the Australian palaeoanthropologist Peter Brown teasingly e-mailed me pictures of a strange-looking skull, asking what I thought it was. I knew that he had been working in east Asia, so I guessed that the images might represent the first discovery of a very primitive member of our genus, *Homo*, from somewhere like China.

Gradually, Brown revealed the even more astonishing news of the skull's remote location and recent age. That October, he, Mike Morwood and colleagues published analyses

in this journal<sup>1,2</sup> with the controversial proposal that the tiny skull and its associated skeleton represented a new human species. They named it *Homo floresiensis*, which Morwood dubbed 'hobbit', owing to its diminutive stature — a moniker that the global press quickly ran with.

The researchers posited that a primitive hominin persisted into the era of anatomically modern humans<sup>2</sup>

and lived in Flores, part of the remote string of Wallacean islands east of Java that have remained isolated since their formation (see 'How did the hobbit get to Flores?'). Controversy about this species continues to this day, including whether it even belongs in *Homo*.

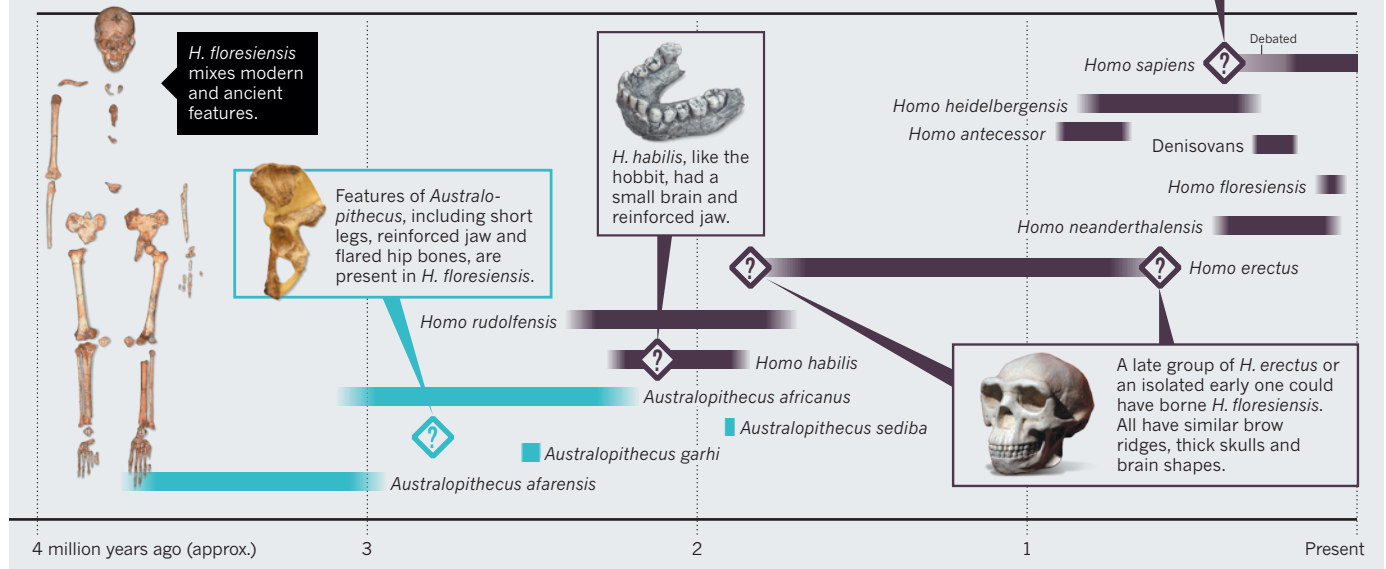
### UNEXPECTED TRIP

In 2004, like most palaeoanthropologists, I thought that only modern humans (*Homo sapiens* — like us) had travelled beyond southeast Asia in the past 60,000 years. ►

► **NATURE.COM**  
See Nature's online special on the hobbit:  
[nature.com/hobbit10](http://nature.com/hobbit10)

## WHERE DOES THE HOBBIT BELONG?

More than a decade after scientists unearthed a startling tiny skull, debate rages over which branch of the human tree bore *Homo floresiensis*.



► By then, people had devised sea-going watercraft essential for such a journey. It seemed unlikely that more-ancient humans could have made such a voyage<sup>3</sup>.

The excavations that first led to the idea that ancient humans did so began in 2001. Morwood, a New Zealand-born archaeologist, led an international team in the huge Liang Bua (meaning 'cool cave') on Flores, hoping to find evidence of the earliest modern humans to colonize Wallacea, Australia and New Guinea. The project reopened trenches several metres deep from previous Dutch and Indonesian work. It soon yielded promising finds: stone tools that seemed to be more than 10,000 years old, and fossils of a pygmy form of the extinct elephant-like *Stegodon*.

In 2003, at a depth of around 6 metres, the team encountered a small skeleton (LB1) that they first thought must represent a modern human child. Then they noticed other details: the wisdom teeth in its jaws had fully erupted, and the tiny skull showed definite brow ridges above its large eye sockets.

The skeleton was dated from associated materials to less than 20,000 years old. Morwood and colleagues argued that it represented a unique example of insular dwarfism in humans. This is a well-known process whereby large mammals isolated on islands evolve smaller bodies in response to limited resources and the lack of predators<sup>4</sup>.

Morwood and colleagues argued that a population of *Homo erectus* could have travelled, perhaps by boat, to Flores from Java (500 kilometres away), where *H. erectus* was first identified in the 1890s. Java, having been repeatedly connected to the rest of Asia over the past 2 million years when sea levels were low, was thought to mark the farthest

extent from Africa of colonization by ancient humans. Morwood and colleagues posited that Flores's ancient settlers underwent island dwarfing, in parallel with other colonizing mammals such as *Stegodon*. Stone tools associated with *Stegodon* bones in Liang Bua suggested that *H. floresiensis* could have hunted and butchered these animals.

### ONGOING ARGUMENTS

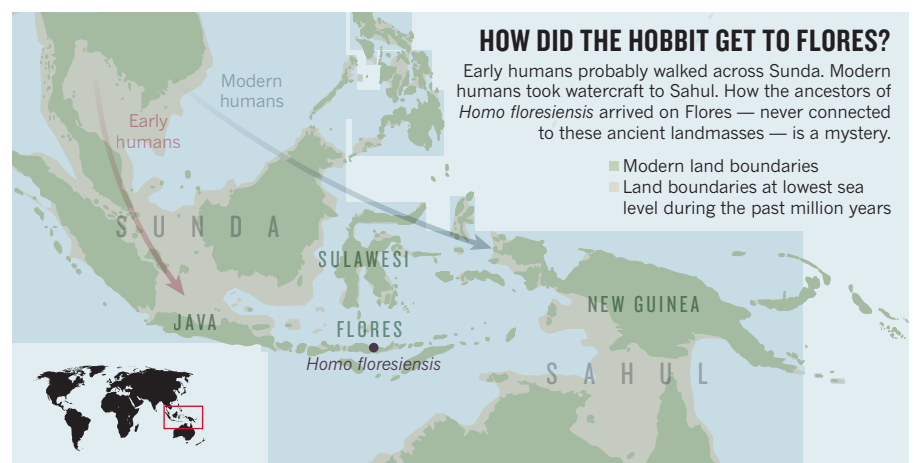
Stone tools discovered elsewhere on Flores, analyses of which were published<sup>5</sup> in 2010, suggest that potential ancestors of *H. floresiensis* could have been on the island a million years ago. But considering an island on the other side of the world — Britain — with its discontinuous record of human settlement over 900,000 years, I can also imagine episodic human colonizations on Flores.

In 2009, a collection of studies<sup>6</sup> analysed LB1 in more detail, along with other fossils attributed to *H. floresiensis*, including

a second jawbone (LB6), and fragments of limb bones of up to eight more individuals. Features such as LB1's broad, flared hipbones, short collarbone, and forwardly positioned shoulder joint all resembled the pre-human group known as australopithecines ('southern apes'), which includes individuals such as the 3.2-million-year-old skeleton of 'Lucy', comparable in size to LB1.

These studies did not settle ongoing arguments about whether the finds represented a small, early human (a *H. erectus* shrunk through insular dwarfing) or an abnormal modern one, wrongly dated and analysed<sup>3,4</sup>. There were further problems: in late 2004, Teuku Jacob, a now-deceased Indonesian palaeoanthropologist, appropriated the specimens to conduct his own work in Yogyakarta. By the time the fossils were returned to Jakarta, following international pressure, some had been damaged irreparably<sup>4</sup>.

The small brain of *H. floresiensis* has





provoked particularly fierce controversy. Some, citing parallels in other dwarfed mammalian species<sup>7,8</sup>, argue that it could derive from a *H. erectus* template, diminished but human in structural organization. Others rule out dwarfing, insisting that the braincase is much smaller than would be expected if a *H. erectus* body were scaled down. They argue that the shape of the brain reflects pathology — perhaps a condition called microcephaly<sup>9</sup>.

Various pathologies can explain some of the unusual aspects of the LB1 skeleton. But in my view, no syndrome so far proposed can account for the totality of evidence from Liang Bua. Neither cretinism, Laron syndrome nor Down's syndrome duplicate the full suite of features.

### CLASSIFYING THE HOBBIT

From the beginning, Brown and Morwood were torn over how to classify the fossil. In the first drafts of their paper they even created an entirely new genus for LB1 to reflect its unique combination of human and non-human traits — '*Sundanthropus floresianus*'. But in the face of insistent reviewers, they shifted to the idea that their find was a dwarfed version of *H. erectus*<sup>4</sup>.

Both Morwood and Brown indicated later that they were not convinced by that model<sup>6,10</sup>, and I join them in their doubts. The tiny brain of LB1, its body shape, and its foot, hand and wrist bones look more primitive than those of any human dating to within the past million years. Primitive traits of the wrist bones and jaw are replicated in at least one more individual from the site<sup>10,11</sup>. Like LB1, the LB6 lower jaw is small, lacks a chin, and shows internal bony reinforcements most like those in pre-human fossils more than 2 million years old<sup>10</sup>.

Ten years on, it is still very difficult to decide between competing views on where the hobbit came from (see 'Where does the hobbit belong?'). Island dwarfing from a local *H. erectus* population is probably still the most widely accepted idea, although this would require the re-emergence of primitive traits as well as convergence on *H. sapiens* in features such as tooth size and shape<sup>12</sup>.

A more primitive origin, from a more ancient *H. erectus* population (such as the 1.8-million-year-old fossils found at Dmanisi in Georgia) would require less extreme dwarfing, but would still need the re-emergence of primitive traits. An even more primitive template, closer to *Homo habilis* or the pre-human australopithecines, is a closer match for the reinforced jawbone, brain and body size, wrist morphology, and body shape, but would require still more convergences on later *Homo* morphology in features such as cranial thickness, retracted face and dental reduction.

We need more bones from Liang Bua to establish the morphological



Liang Bua cave on the Indonesian island of Flores, the discovery site of *Homo floresiensis*.

variation of *H. floresiensis* and set pathological explanations to rest. At present we do not even know the extent of sexual dimorphism in the species — would a male skeleton be much larger and more *H. erectus*-like?

Isotope studies and analyses of preserved dental tartar could help to reconstruct the *H. floresiensis* diet, and investigations of dental microstructure might place the species taxonomically, because primitive hominins grew distinctly faster than *H. erectus* and later humans<sup>3</sup>. Even small amounts of ancient DNA would greatly clarify its evolutionary history, but it will require both technological breakthroughs and good fortune to acquire analysable samples from the warm, wet conditions of Liang Bua.

Significant work on re-evaluating the dates of the site, fossils and archaeology was under way before Morwood's untimely death in 2013. The results, due soon, will undoubtedly affect our views of *H. floresiensis*, and when and why it went extinct.

### MORE SURPRISES

I think that there are more surprises to come from the rest of Wallacea. If the ancestors of *H. floresiensis* reached Flores, perhaps they also dispersed to other islands, and the experiment in human evolution revealed in Liang Bua might have equally remarkable parallels elsewhere — for example on Sulawesi, the Philippines and Timor. As Morwood pointed out<sup>4,6</sup>, the powerful currents around Indonesia would have favoured transport from Sulawesi (north of Flores) rather than from Java, where the nearest *H. erectus* fossils have been found. The possibility of accidental rafting on mats of vegetation in such a tectonically active region must also be considered;

in the 2004 Indian Ocean tsunami, some people who survived on floating debris were dispersed more than 150 kilometres.

If the *H. floresiensis* lineage had a more primitive origin than the oldest known *H. erectus* fossils so far identified in Asia, then we would have to re-evaluate the dominant explanation for how humans arose and spread from Africa. Most current thinking assumes that the first dispersal from Africa was just before the time of the Dmanisi fossils<sup>3</sup>. An ancient origin for the hobbit would make that dispersal earlier and more complex<sup>13</sup>. It would mean that a whole branch of the human evolutionary tree in Asia had been missing until those fateful discoveries in Liang Bua. ■ **SEE NEWS FEATURE P.422**

**Chris Stringer** is a research leader in human origins at the Natural History Museum in London, UK.  
e-mail: c.stringer@nhm.ac.uk

1. Morwood, M. J. *et al.* *Nature* **431**, 1087–1091 (2004).
2. Brown, P. *et al.* *Nature* **431**, 1055–1061 (2004).
3. Stringer, C. *The Origin of Our Species* (Penguin, 2012).
4. Morwood, M. & van Oosterzee, P. *A New Human: The Startling Discovery and Strange Story of the "Hobbits" of Flores, Indonesia* (Harper Collins, 2000).
5. Brumm, A. *et al.* *Nature* **464**, 748–752 (2010).
6. Morwood, M. J. & Jungers, W. L. *J. Hum. Evol.* **57**, 640–648 (2009).
7. Falk, D. *et al.* *J. Hum. Evol.* **57**, 597–607 (2009).
8. Kubo, D., Kono, R. T. & Kaifu, Y. *Proc. R. Soc. B* **280**, 20130338 (2013).
9. Vannucci, R. C., Barron, T. F. & Holloway, R. L. *Anat. Rec.* **296**, 630–637 (2013).
10. Brown, P. *J. Hum. Evol.* **62**, 201–224 (2012).
11. Orr, C. M. *et al.* *J. Hum. Evol.* **64**, 109–129 (2013).
12. Baab, K. L., McNulty, K. P. & Harvati, K. *PLoS ONE* **8**, e69119 (2013).
13. Dennell, R. & Roebroeks, W. *Nature* **438**, 1099–1104 (2005).



# Be prepared

Scenario-based training for disasters is better than just drawing up a paper plan, say **Jennifer K. Pullium** and colleagues.

Two years after Hurricane Sandy hit the US east coast, we are often asked what we learned from the disaster. Some lessons sound trivial, such as the importance of keeping headlamps ready and accessible. Some are worth stressing: we learned that we had a good contingency plan. But the most important thing we learned is that the way people act during a disaster can be more important than the written plan.

We are veterinary surgeons and animal-laboratory managers at the New York University (NYU) Langone Medical Center. Prior to Sandy, our emergency preparation comprised lectures and exercises. This served us well, but our contingency plan did not anticipate staff who care for lab animals having to break through a ceiling and lower a basket to trained hazard-response professionals to rescue laboratory mice. Ten days into the disaster, our team was stressed, cold and exhausted. We triaged, treated and transported mice for seven straight hours until late into the night. We rescued 600 cages with thousands of mice, many of them unique strains that investigators feared were lost entirely<sup>1</sup>.

Standard emergency preparations do not account for taxed and terrified minds, and tendencies to make poor decisions. The strength of the people in our team came from skills unrelated to their day jobs. But previous experiences came in handy — one of us (M.A.R.) served in the US Army, and another of us (J.K.P.) is a licensed pilot. Such training enabled us to make good decisions under pressure. The US military has trained soldiers for appropriate responses to stressful conditions for decades. Pilot instruction teaches people to efficiently use complex, systematic approaches in times of crisis to assess risks, instruct crews and manage stress.

These kinds of skills are not taught in typical disaster training<sup>2</sup>. They should be. In the months after the flood waters ebbed, we created a series of tactical decision games to improve employees' abilities to lead responses and to assess and communicate situations. These go far beyond paper plans.

## SIMULATED STRESS

Institutions typically develop plans while sitting around a conference table in a comfortable room with refreshments. Practice sessions are run with the entire staff on a warm, sunny day. They do not prepare people to act decisively. When surveyed, most animal-resource staff at our facility said that they



A steel door pushed in by Hurricane Sandy at the NYU Langone Medical Center.

would attempt to contact upper management before doing anything in an emergency. But if leaders are absent during crises, indecision could cause destructive delays. The best-written disaster plan is not worth the paper if it is printed on without people on site able to execute it.

Training can shape contingency plans to account for particular personalities. During our emergency response, we saw clear aptitudes in two people who were employed as cage-washers, an entry-level position. In the days following Sandy, these two never hesitated. They invigorated and motivated their teams, and stayed positive and focused. Leading by example, they encouraged higher-ranking personnel to keep working, even when tired and hungry. They were integral members of our impromptu rescue team, evaluating the situation and making decisions. They determined the route to transport rescued animals to another NYU Langone facility, a trek that spanned several floors and buildings without power. Had we known of their capabilities earlier, we might have made them leaders of small teams from the start of the emergency.

Training can empower personnel to feel competent to step into leadership voids. Too many training programmes look for responses guided by standard operating procedures, with predetermined 'correct' answers. By contrast, tactical decision games

simulate stressful, challenging situations and require participants to make choices without full information or clearly correct answers<sup>2</sup>. Exercises that were initially developed for military use have been adapted for civilians by industrial psychologists such as Margaret Crichton. Such training is becoming standard in aviation, nuclear-power plants and medicine<sup>3</sup>. After coping with Sandy, our team created a programme for animal-care facilities.

## DECISION TIME

Our exercises put trainees on their mettle. We allow too little time for people to assess disaster scenarios. Instead of analysing the situation, we demand that trainees state their next actions; we force them to make decisions on the spot. When a trainee says that he would call for help, we hand over a phone to hear what he would say. Then we ask what he will do when emergency responders or leaders reply that they are not coming. Everyone gets a turn in the 'hot seat', with individuals scrutinized both by the trainer and by their fellow participants. In our classes, distractions such as air horns, darkness and flashing lights increase stress.

People may not enjoy these exercises, but they do see the value. Working through shifting scenarios allows trainees to become confident in their thought processes and abilities. When people are reduced to stumbling responses, this is framed as an area to improve. In our exercises, much as in a real disaster, a right answer is not essential; making a decision is.

Such training can be easily incorporated into other team-building exercises in a busy work environment. We conduct disaster training annually. Staff learn crucial lessons about how to lead in a crisis. Even if the skills developed are never put to use in a real disaster, facilities, people and animals are still likely to benefit. ■

**Jennifer K. Pullium** is director and attending veterinarian, **Gordon S. Roble** is veterinarian, and **Mark A. Raymond** is associate director in the Division of Laboratory Animal Resources, New York University Langone Medical Center, New York, USA.

e-mail: [jennifer.pullium@nyumc.org](mailto:jennifer.pullium@nyumc.org)

1. Fishell, G. *Nature* **496**, 421–422 (2013).
2. Crichton, M. *Horiz. Psychol.* **10**, 7–22 (2001).
3. Crichton, M. T. & Flin, R. *Ann. Nucl. Energy* **31**, 1317–1330 (2004).



Mary Somerville helped to unify science intellectually at a time of specialization.

## IN RETROSPECT

# On the Connexion of the Physical Sciences

**Richard Holmes** finds Mary Somerville's breakthrough science best-seller thrillingly fresh, 180 years on.

Less than two centuries ago, popular science barely existed. In 1830, astronomer John Herschel wrote to natural philosopher William Whewell about the urgent need for "digests of what is actually known in each particular branch of science ... to give a connected view of what has been done, and what remains to be accomplished".

The remarkable writer who first achieved that "connected view" and arguably launched popular science writing was a self-taught Scottish mathematician, Mary Fairfax Somerville (1780–1872). Her book, *On the Connexion of the Physical Sciences*, published by John Murray in 1834 alongside works by Walter Scott, Lord Byron and Jane Austen, contains no equations,

few diagrams and little mathematics. But it is a masterpiece of descriptive explanation and analogy that unveils a complete scientific world view, covering everything from stars to insects. It was Murray's best-selling scientific publication until Charles Darwin's *On the Origin of Species* in 1859; it eventually ran to ten editions in Britain, and was published in France, Italy, Germany and the United States.

The book appeared at a critical moment for science. The disciplines were beginning to define their territories and societies were starting to coalesce, including the Geological Society in 1807 and the Zoological Society of London in 1826. For Whewell, who scrutinized Somerville's offering in depth in the *Quarterly Review* of March 1834, the work was a "masterly" survey that performed the crucial task of intellectual unification at a moment when science threatened to become like "a great empire falling to pieces".

Somerville, the daughter of admiral William Fairfax, grew up near Scotland's Firth of Forth wandering the seashore, collecting shells and studying seabirds. Her father described her as a savage. At 15 she saw a mysterious reference to algebra in a women's fashion magazine, and began to devour the theorems of Euclid and Newton. She would lie in bed at night, listening to the sea and solving equations in her head.

She was launched into Edinburgh society at 18, a notable beauty; to her parents' dismay she continued her study of maths, as well as painting and the piano, describing herself as "intensely ambitious to excel in something, for I felt in my own breast that women were capable of taking a higher place in creation than that assigned to them in my early days". A disastrous first marriage to glamorous naval attaché Samuel Greig was cut short by his early death in 1807.

In 1812 she met and married her soulmate and cousin, the scientifically minded, globe-trotting physician William Somerville. Moving to London in 1816, and now the parents of four children, the couple entertained leading scientists such as William Herschel, Michael Faraday, Charles Babbage and Charles Lyell. In 1830, at the invitation of the Society for the Diffusion of Useful Knowledge, she translated French astronomer Pierre-Simon Laplace's highly technical *The Mechanism of the Heavens*. In a parliamentary debate on scientific education, she was referred to as "one of the only six persons in England who understands Laplace".

Somerville began writing *On the Connexion of the Physical Sciences* in 1832,

**On the Connexion of the Physical Sciences**

MARY SOMERVILLE

John Murray: 1834.

THOMAS PHILLIPS/MARY FAIRFAX; MRS WILLIAM SOMERVILLE/SCOTTISH NATL PORTRAIT GALLERY



during a long visit to Paris. She effectively became an expert reporter on the latest developments in European science. Taking full advantage of social networking, she contacted Laplace's influential widow and dined with the physicists François Arago, Jean-Baptiste Biot and Joseph-Louis Gay-Lussac. She had privileged status at sites from the Paris Observatory to the National Museum of Natural History, and in the laboratories of electrical-theory pioneers André-Marie Ampère and Antoine César Becquerel.

In contrast to the vague speculations of eighteenth-century natural philosophy, her 500-page book covers a tight field of hard sciences — astronomy, physics, chemistry, geography, meteorology and electromagnetism. Its groundbreaking style, clear and logical, occasionally opens out into passages of sublime perspective, such as the description of universal gravity as a force equally present “in the descent of a rain drop as in the falls of Niagara; in the weight of the air, as in the periods of the moon”. Somerville ranges over subjects from stellar parallax to terrestrial magnetism, from comets to giant seaweed.

Her handling of acoustics is characteristically brilliant, based on the observations of John Herschel, Arago and naturalist Alexander von Humboldt. Comparing the propagation of sound to “a field of corn agitated by a gust of wind”, she goes on to describe phenomena from birdsong to thunder. She also suggests a connection between waves propagated in water, the atmosphere and sunlight, writing: “Any one who has observed the reflection of the waves from a wall on the side of a river ... after the passage of a steam-boat, will have a perfect idea of the reflection of sound and of light.”

Her exploration of the solar spectrum contains one of the earliest descriptions (derived from work by William Herschel, chemist William Hyde Wollaston and physicist Johann Wilhelm Ritter) of infrared and ultraviolet rays at the extreme ends of the known light spectrum, “too extensive in their undulations to affect our optic nerves”. She speculates that such rays might have many possible functions in the animal kingdom: “We are altogether ignorant of the perceptions which direct the

carrier-pigeon to his home ... or of those in the antennae of insects which warn them of the approach of danger”. She also mused about climate change, the cause of earthquakes and the existence of planets beyond Uranus.

(pictured), the geographical explorations of Lyell and Humboldt, and the teams of European astronomers who observed the return of Halley's comet, among other feats.

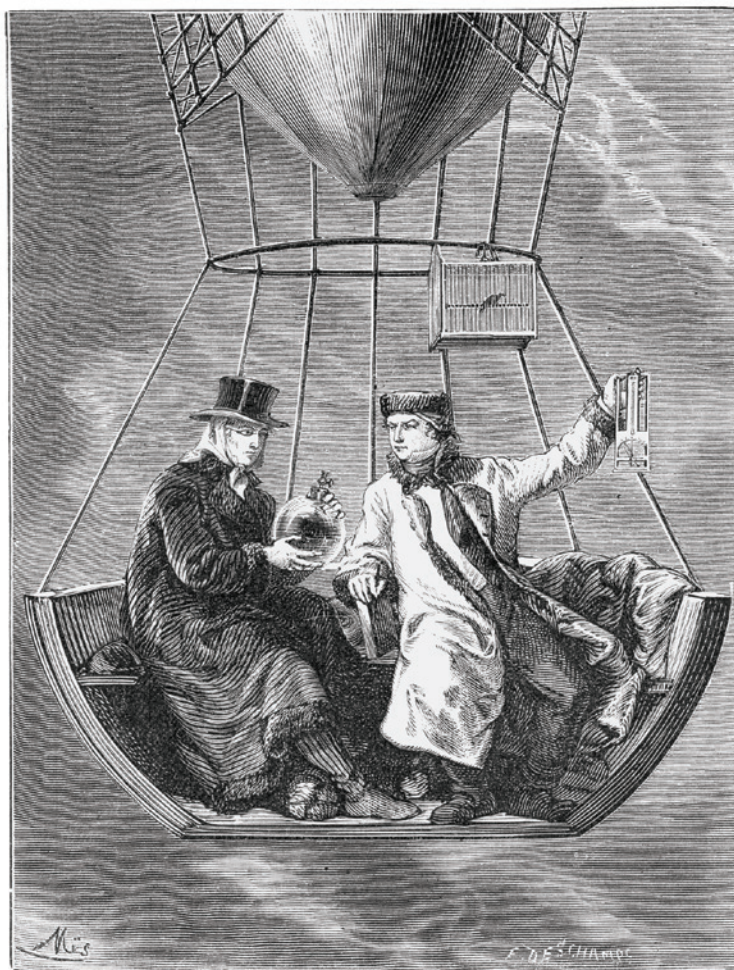
*On the Connexion of the Physical Sciences* was widely praised by journalists and scientists in Britain and abroad; both Arago and Humboldt deeply admired it. The popular, large-circulation journal *Mechanics' Magazine* urged its audience: “read it! read it!” Somerville dined at the male stronghold of the University of Cambridge, invited by its science professors; received honorary membership of the Royal Astronomical Society among others; and, although barred from the Royal Society, is commemorated there in a formidable marble bust.

Such was her celebrity that she wrote, “I am a kind of tame Lioness at present.” Her friend the novelist Maria Edgeworth, however, noted that “while her head is up among the stars, her feet are firm upon the earth”. Somerville, privately unorthodox and witty, sceptical but still believing in a creator, lived up to that estimation throughout her long life. She supported women's suffrage (her signature was the first on philosopher John Stuart Mill's petition to Parliament); campaigned against vivisection, and against slavery in America; and believed in Darwinian evolution.

Like the great poets of her era, Somerville brought a new vision into the world, and one that a broad, educated public could grasp. Seven years after her death, a new women's college at the University of Oxford was named in her honour.

Looking back almost 40 years after the publication of her magnum opus, Maxwell reflected: “It was one of those suggestive books which put into definitive, intelligible and communicative form the guiding ideas that are already working in the minds of men of science”. In fact, the book prompted the creation of a new professional concept, and a new umbrella word to define it, coined by Whewell in his review of 1834: “scientist”. ■

**Richard Holmes** is the author of *The Age of Wonder*, which won the 2009 Royal Society Prize for Science Books, and *Falling Upwards: How We Took to the Air*. e-mail: richard.holmes@osb.me.uk



French scientists making high-altitude measurements in 1804.

The most original sections deal with electricity and the new science of electromagnetism. Somerville thrillingly describes Faraday's latest work with the horseshoe magnetic generator, establishing that magnetism and electricity must have complex links in what he was beginning to define as ‘fields’. These sections clearly predict the connection between all electromagnetic phenomena, established a generation later by physicist James Clerk Maxwell (see [nature.com/maxwell](http://nature.com/maxwell)).

Somerville's work contextualized the sciences as an ongoing global project. The book emphasized, in a wholly new way, the communal nature of science as shared discovery, referring to John Franklin's Arctic expeditions, the high-altitude balloon flights of Biot and Gay-Lussac

➔ **NATURE.COM**  
For Richard Holmes on science biography, see: [go.nature.com/9atgoi](http://go.nature.com/9atgoi)



# Correspondence

## Warming goal: still the best indicator

David Victor and Charles Kennel challenge the practice of using global mean temperature as the main measure of danger from climate change (*Nature* **514**, 30–31; 2014). On the basis of 40 years of science and policy research, there are good reasons why this temperature is the favoured indicator.

It can be related through climate models to the regional impacts and risks that drive public concern (see [go.nature.com/5chktj](http://go.nature.com/5chktj)). It is indeed “related only probabilistically to emissions”, but the authors’ best indicator — carbon dioxide concentration — is related only probabilistically to impacts and risks, except in the case of ocean acidification. As for ocean heat content, its trend experiences interruptions much like the global mean temperature, and bears no direct relationship to most impacts and risks.

Compared with other proposals, global mean temperature is more closely related to outcomes for people and ecosystems. Without such a goal, we shall never know how much reduction in emissions is sufficient.

**Michael Oppenheimer** Princeton University, New Jersey, USA.  
[omichael@princeton.edu](mailto:omichael@princeton.edu)

## Warming goal: clear link to emissions

David Victor and Charles Kennel argue that aiming to keep average global warming within 2 °C of pre-industrial temperatures is neither politically nor scientifically useful (*Nature* **514**, 30–31; 2014). I disagree: global temperature change is the closest thing we have to a metric with a clear link to emissions; it can also be related quantitatively to a range of local climate impacts.

Because global temperature seems to respond linearly to cumulative emissions of carbon

dioxide (H. D. Matthews *et al.* *Nature* **459**, 829–832; 2009), policies to cut emissions should also reduce global temperature change. This offers a simple framework for estimating a global carbon budget that contains warming to within 2 °C.

Policy goals should not have adverse effects on human and environmental welfare. Using global temperature avoids these too, because it seems to be an indicator of the extent of local climate changes (see, for example, M. Markovic *et al.* *Clim. Change* **120**, 197–210; 2013). Furthermore, the average global temperature over decades relates well to many climate impacts and to Victor and Kennel’s ‘vital signs’ of planetary health (National Research Council *Climate Stabilization Targets* National Academies Press, 2011).

Now that the international community has finally coalesced around the 2 °C goal, compelling reasons are needed to interrupt this momentum.

**H. Damon Matthews** Concordia University, Montreal, Canada.  
[damon.matthews@concordia.ca](mailto:damon.matthews@concordia.ca)

## Open access to Earth land–cover map

China last month donated to the United Nations the first open-access, high-resolution map of Earth’s land cover, as a contribution towards global sustainable development and combating climate change.

The map, known as GlobeLand30, comprises data sets collected at 30-metre resolution — more than ten times that of previous data sets. These data sets will be valuable for monitoring environmental changes and for resource management at global, regional and local scales (see also M. A. Wulder and N. C. Coops *Nature* **513**, 30–31; 2014).

The GlobeLand30 data sets are freely available and comprise ten types of land cover, including forests, artificial surfaces and wetlands, for the years 2000 and

2010. They were extracted from more than 20,000 Landsat and Chinese HJ-1 satellite images (see [www.globallandcover.com](http://www.globallandcover.com)).

GlobeLand30 will promote scientific data sharing in the fields of Earth observation and geospatial sciences.

**Chen Jun** National Geomatics Center of China, Beijing, China.  
[chenjun@nsdi.gov.cn](mailto:chenjun@nsdi.gov.cn)

**Yifang Ban** KTH Royal Institute of Technology, Stockholm, Sweden.  
**Songnian Li** Ryerson University, Toronto, Ontario, Canada.

## Sustainability: root targets in consensus

Mark Stafford-Smith urges scientists to engage more effectively with the United Nations’ Sustainable Development Goals to ensure that their environmental targets are quantifiable (*Nature* **513**, 281; 2014). When stakeholder values are diverse and passionately defended, however, such targets may not be easily agreed — leading to stalled negotiations and stagnant progress on issues of global significance.

In our view, building consensus over desirable environmental outcomes would be a better approach. This involves analysing different possible outcomes, understanding decision-making processes and improving communication among stakeholders who have conflicting interests.

Initiatives such as Future Earth and the Intergovernmental Platform on Biodiversity and Ecosystem Services are helping scientists to engage with international environmental policy. In the ongoing negotiations over the Sustainable Development Goals, scientists need to move on from simple information provision and help to develop appropriate policies.

**Sean Maxwell\*** University of Queensland, Brisbane, Australia.  
[smaxwell@uq.edu.au](mailto:smaxwell@uq.edu.au)

\*On behalf of 10 correspondents (see [go.nature.com/tqxjyj](http://go.nature.com/tqxjyj) for full list).

## Stop the cuts, not the evaluations

Amaya Moro-Martin asserts that the European Science Foundation (ESF) supported a “flawed evaluation process” for research in Portugal (*Nature* **514**, 141; 2014). This unsubstantiated allegation undermines the foundation’s work and is detrimental to the many excellent reviewers and panel members involved in the evaluation process.

The ESF champions the benefits to society from investments in research. We are very concerned about the increased pressure on many national science budgets. However, we believe that peer review, despite its limitations, is the most meritocratic and evidence-based approach to resource allocation. The work of those public-spirited scientists willing to give their time and energy to the peer-review process must be acknowledged, respected and supported. They should be allowed to undertake their work without interference.

During the course of the independent research evaluation implemented for the Foundation for Science and Technology in Portugal, the ESF has witnessed an unprecedented level of direct interference with peers and panel members in the performance of their work. Even while the review process is ongoing, many have received intimidating communications designed to discourage them from completing their agreed tasks. This practice is unacceptable and damaging to science.

It is in this context that we respond to Moro-Martin’s remark. Although no evaluation process is perfect, it is the most independent system yet devised. The ESF has carried out this evaluation project in accordance with good practice (see [go.nature.com/o4xfuz](http://go.nature.com/o4xfuz); to be updated on completion of the project).

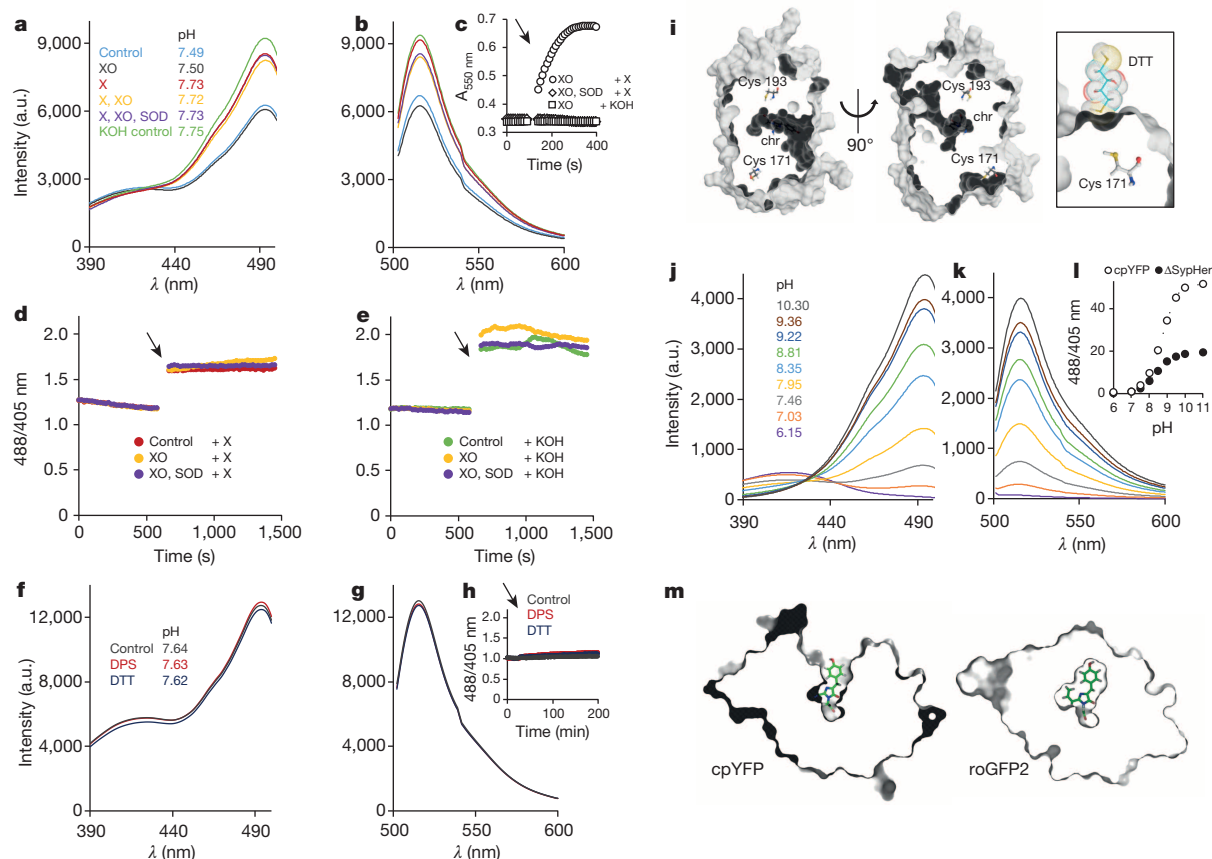
**Jean-Claude Worms**, **Jane Swift** ESF, Strasbourg, France.  
[jswift@esf.org](mailto:jswift@esf.org)

# The ‘mitoflash’ probe cpYFP does not respond to superoxide

ARISING FROM E.-Z. Shen *et al.* *Nature* **508**, 128–132 (2014); doi:10.1038/nature13012

Ageing and lifespan of organisms are determined by complicated interactions between their genetics and the environment, but the cellular mechanisms remain controversial; several studies suggest that cellular energy metabolism and free radical dynamics affect lifespan, implicating mitochondrial function. Recently, Shen *et al.*<sup>1</sup> provided mechanistic insight by reporting that mitochondrial oscillations of ‘free radical production’, called ‘mitoflashes’, in the pharynx of three-day old *Caenorhabditis elegans* correlated inversely with lifespan. The interpretation of

mitoflashes as ‘bursts of superoxide radicals’ assumes that circularly permuted yellow fluorescent protein (cpYFP) is a reliable indicator of mitochondrial superoxide<sup>2</sup>, but this interpretation has been criticized because experiments and theoretical considerations both show that changes in cpYFP fluorescence are due to alterations in pH, not superoxide<sup>3–7</sup>. Here we show that purified cpYFP is completely unresponsive to superoxide, and that mitoflashes do not reflect superoxide generation or provide a link between mitochondrial free radical dynamics and lifespan. There is



**Figure 1 | Spectroscopic and structural analysis of the ‘mitoflash’ probe cpYFP.** **a**, **b**, cpYFP fluorescence excitation spectra (emission at 515 nm) (**a**) and emission spectra (excitation at 488 nm) (**b**) after addition of xanthine (X; 2 mM), xanthine oxidase (XO; 100 mU ml<sup>-1</sup>), bovine Cu/Zn superoxide dismutase (SOD; 600 U ml<sup>-1</sup>) and KOH (solvent control for xanthine; note pH increase), and the *in situ* pH of the assayed 200 μl reaction mix. a.u., arbitrary units. **c**, Cytochrome *c* reduction detected by absorption at 550 nm to measure superoxide generation in response to the X/XO system in the presence and absence of SOD, and in response to KOH as a solvent control for X. Controls contained either SOD addition or no XO. **d**, The response of cpYFP excitation ratio (488/405 nm) to superoxide generation. Arrow indicates introduction of X to constitute the X/XO system. Controls contained either SOD addition or no XO. **e**, The same assays were performed with KOH introduction as solvent control for X. **f**, **g**, cpYFP fluorescence excitation spectra (**f**) and emission spectra (**g**) after 24 h incubation with dithiothreitol (DTT; 10 mM) and 2,2'-dipyridyl disulphide (DPS; 1 mM), and the *in situ* pH

of the assayed 200 μl reaction mix. **h**, cpYFP fluorescence ratio after DTT and DPS addition over 3 h; arrow indicates DTT or DPS addition. **i**, Cross-sections through a surface model of cpYFP (left and middle). Chromophore (chr) and cysteine residues are represented as ball-and-stick models. The Cys 171 thiol is relatively close to the protein surface, but unlikely to be accessible to solutes as indicated by the docking of a DTT molecule to the protein surface (right). **j**, **k**, cpYFP fluorescence excitation spectra (emission at 515 nm) (**j**) and emission spectra (excitation at 488 nm) (**k**) in response to pH as determined *in situ* after the measurements. **l**, pH dependence of cpYFP excitation ratio (488/405 nm; normalized to pH 7.0) as compared to the cpYFP part of the pH sensor SypHer (ΔSypHer). **m**, Sectional views through volume models of cpYFP and redox-sensitive GFP (roGFP2), with the clipping plane parallel to and just above the chromophore phenoxy ring. The chromophore is represented as a ball-and-stick model. Data in **a–h** and **j–l** are background corrected, and experiments were repeated at least five times with consistent results.



a Reply to this Brief Communication Arising by Cheng, H. *et al.* *Nature* **514**, <http://dx.doi.org/10.1038/nature13859> (2014).

We carried out experiments with purified recombinant cpYFP sensor protein to test whether it responds to superoxide (Fig. 1a–e). Exposure of cpYFP to a superoxide-generating system (xanthine (X) and xanthine oxidase (XO)) slightly changed the excitation and emission spectra. However, the same change occurred when cpYFP was incubated with the individual assay constituents in the absence of superoxide production, or when Cu/Zn superoxide dismutase (SOD) was added to degrade superoxide (Fig. 1a, b). The cytochrome *c* reduction assay confirmed that superoxide is produced by the X/XO system, and is abolished by SOD (Fig. 1c). Xanthine is dissolved in potassium hydroxide, causing a small increase in pH after addition. There was an excellent correlation between spectral changes and resulting assay pH after xanthine (that is, potassium hydroxide) addition (Fig. 1a, b). In time course assays in which superoxide generation was started by the addition of xanthine (Fig. 1d), the addition of potassium hydroxide as the solvent control for xanthine (Fig. 1e) gave the same increase in fluorescence ratio. Extended reductive or oxidative treatment with thiol redox agents (the reducing agent dithiothreitol (DTT) and oxidizing agent 2,2'-dipyridyl disulphide (DPS)) did not alter the spectral behaviour (Fig. 1f–h), consistent with structural information suggesting that both Cys residues are buried inside the mature protein and are unlikely to be accessible for thiol redox chemistry (Fig. 1i). Likewise, reductive pre-treatment with DTT under inert atmosphere, followed by DTT removal, did not affect the outcome of the superoxide assays. Further variation of experimental variables, including pre-incubation conditions, pH buffer systems and a 100-fold range of sensor concentrations, did not lead to any rapid, reversible change in cpYFP sensor signal required for superoxide-related mitoflashes, as long as the pH and halide ion concentrations were kept constant.

Mitoflashes can be fully explained by the extraordinary pH sensitivity of cpYFP, which has a  $pK_a$  value of  $\sim 8.7$  (determined by measuring fluorescence after excitation at 488 nm, the wavelength at which flashes are observed) and shows a  $>50$ -fold change in fluorescence ratio between pH 7 and 10, similar to the structurally related ratiometric pH-sensor SypHer<sup>8</sup> (Fig. 1j–l). In the mitochondrial matrix, a resting pH ( $\sim 7.9$ ) close to sensor  $pK_a$  and a limited pH buffering capacity mean that even minor perturbations will elicit a pronounced sensor response (Fig. 1a, b, d, e). The cpYFP pH sensitivity is due to the structural perturbation caused by the circular permutation. A large cleft in the  $\beta$ -barrel exposes the pH-active phenoxy group of the chromophore (Fig. 1m, left), which is concealed in non-permuted green fluorescent protein (GFP)-based biosensors (Fig. 1m, right).

On the basis of this evidence using purified cpYFP and earlier studies in cells and isolated mitochondria<sup>5,6,9</sup>, the mitoflash phenomenon cannot be attributed to bursts of mitochondrial superoxide. In accordance with the pH responsiveness of the probe, recent work with different sensors suggests that mitoflash events indicate brief periods of alkalisation in individual mitochondria, possibly as a result of acceleration in proton pumping, triggered by mitochondrial fusion initiation and/or a change in ion homeostasis<sup>6,9,10</sup>.

The debate about the nature of mitoflashes has focused on *in situ* evidence that has left space for interpretation on both sides. Critics have pointed out the implausibility of 'superoxide flashes' on the basis of mitochondrial energetics<sup>3,5,9</sup>, the absence of a plausible chemical mechanism for the reversible interaction between cpYFP and superoxide<sup>4,7</sup>, and the fact that the pH sensor SypHer also detects mitoflashes<sup>6,10</sup>. These arguments have been countered by data suggesting a correlation of mitoflashes with the response of chemical probes for reactive oxygen species<sup>11–13</sup>, the notion that the pH probe SypHer may also respond to superoxide<sup>14</sup>, and the suggestion that a mitoflash represents a mixture of superoxide burst and pH transient<sup>11,13</sup>. Ultimate resolution of the debate has been hampered by the use of different biological systems and the complexity of

mitochondrial physiology, where matrix pH and free radical release are connected by the electron transport chain and linked to several other parameters such as availability of respiratory substrates, membrane potential, redox and ion homeostasis, and mitochondrial morphology<sup>2,5–7,10,15–17</sup>. Here we resolve the controversy by a thorough analysis of the fundamental properties of the mitoflash sensor cpYFP. Previous work already excluded the suggestion that the pH probe SypHer responds to superoxide<sup>6</sup>. We now provide definitive evidence that cpYFP itself does not respond to superoxide and that flashes recorded by cpYFP do not represent superoxide bursts. Of course, sudden changes in mitochondrial physiology may still include altered free radical levels. Although the mitoflash phenomenon may reflect an important feature of mitochondrial function that deserves further mechanistic analysis, the interpretation of the events by Shen *et al.* lacks a biophysical foundation and mitoflashes cannot serve as evidence for free radical involvement in determining lifespan.

## Methods

cpYFP was purified from *Escherichia coli* Origami (DE3) and Rosetta 2 (DE3) 24 h after induction at 20 °C and assayed at 10, 25 and 1,000  $\mu\text{g ml}^{-1}$  using a Jasco spectrofluorimeter FP8300 and a BMG Labtech Clariostar plate reader. Detector gain was adjusted for individual experiments. Buffers contained 100 mM NaCl, 1 mM  $\text{Na}_2\text{EDTA}$  and 100 mM Tris-HCl, pH 7.5 (for thiol redox and superoxide assays; degassed and under argon for thiol redox treatments) or 100 mM Tris-TES (for pH assays). All reagents were dissolved in assay buffer, except for xanthine (100 $\times$  stock in 1 M KOH, base required for solubility) and xanthine oxidase (118 $\times$   $(\text{NH}_4)_2\text{SO}_4$  suspension as delivered by Sigma). Protein structures (PDB entries 3O78 and 1JC1) were rendered using PyMOL.

**Markus Schwarzländer<sup>1</sup>, Stephan Wagner<sup>1</sup>, Yulia G. Ermakova<sup>2</sup>, Vsevolod V. Belousov<sup>2</sup>, Rafael Radi<sup>3</sup>, Joseph S. Beckman<sup>4</sup>, Garry R. Buettner<sup>5</sup>, Nicolas Demaurex<sup>6</sup>, Michael R. Duchen<sup>7</sup>, Henry J. Forman<sup>8,9</sup>, Mark D. Fricker<sup>10</sup>, David Gems<sup>11</sup>, Andrew P. Halestrap<sup>12</sup>, Barry Halliwell<sup>13</sup>, Ursula Jakob<sup>14</sup>, Iain G. Johnston<sup>15</sup>, Nick S. Jones<sup>15</sup>, David C. Logan<sup>16</sup>, Bruce Morgan<sup>17</sup>, Florian L. Müller<sup>18</sup>, David G. Nicholls<sup>19</sup>, S. James Remington<sup>20</sup>, Paul T. Schumacker<sup>21</sup>, Christine C. Winterbourn<sup>22</sup>, Lee J. Sweetlove<sup>10</sup>, Andreas J. Meyer<sup>1</sup>, Tobias P. Dick<sup>17</sup> & Michael P. Murphy<sup>23</sup>**

<sup>1</sup>Institute of Crop Science and Resource Conservation (INRES), University of Bonn, Friedrich-Ebert-Allee 144, 53113 Bonn, Germany. email: markus.schwarzlander@uni-bonn.de

<sup>2</sup>Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, 117997, Russia.

<sup>3</sup>Departamento de Bioquímica, and Center for Free Radical and Biomedical Research, Facultad de Medicina, Universidad de la República, Avda. General Flores 2125, 11800 Montevideo, Uruguay.

<sup>4</sup>Linus Pauling Institute, Environmental Health Sciences Center, Department of Biochemistry and Biophysics, Oregon State University, Corvallis, Oregon 97331, USA.

<sup>5</sup>The University of Iowa, Department of Radiation Oncology and Interdisciplinary Graduate Program in Human Toxicology, and ESR Facility, College of Medicine, Med Labs B180K, Iowa City, Iowa 52242-1181, USA.

<sup>6</sup>Department of Cell Physiology and Metabolism, University of Geneva, 1, rue Michel-Servet, Geneva 4 CH-1211, Switzerland.

<sup>7</sup>Department of Cell and Developmental Biology and Consortium for Mitochondrial Research, University College London, Gower Street, London WC1E 6BT, UK.

<sup>8</sup>Life and Environmental Sciences Unit, University of California, Merced, 5200 North Lake Road, Merced, California 95344, USA.

<sup>9</sup>Andrus Gerontology Center of the Davis School of Gerontology, University of Southern California, 3715 McClintock Avenue, Los Angeles, California 90089-0191, USA.

<sup>10</sup>Department of Plant Sciences, University of Oxford, South Parks Road, Oxford OX1 3RB, UK.

<sup>11</sup>Institute of Healthy Ageing, and Department of Genetics, Evolution and Environment, University College London, London WC1E 6BT, UK.

<sup>12</sup>School of Biochemistry and Bristol CardioVascular, University of Bristol, Medical Sciences Building, University Walk, Bristol BS8 1TD, UK.

<sup>13</sup>Department of Biochemistry, National University of Singapore, Singapore 117597, Singapore.

<sup>14</sup>Molecular, Cellular and Developmental Biology Department, University of Michigan, Ann Arbor, Michigan 48109-1048, USA.

<sup>15</sup>Department of Mathematics, South Kensington Campus, Imperial College London, London SW7 2AZ, UK.

<sup>16</sup>Université d'Angers & INRA & Agrocampus Ouest, UMR 1345 Institut de Recherche en Horticulture et Semences, Angers, F-49045, France.

<sup>17</sup>Division of Redox Regulation, German Cancer Research Center (DKFZ), DKFZ-ZMBH Alliance, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany.

<sup>18</sup>Department of Cancer Biology, University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA.

<sup>19</sup>Buck Institute for Research on Aging, 8001 Redwood Boulevard, Novato, California 94945, USA.

<sup>20</sup>Department of Physics, Institute of Molecular Biology, University of Oregon, Eugene, Oregon 97403-1229, USA.

<sup>21</sup>Department of Pediatrics, Division of Neonatology, Northwestern University, Feinberg School of Medicine, Chicago, Illinois, 60611, USA.

<sup>22</sup>Centre for Free Radical Research, Department of Pathology, University of Otago, Christchurch, PO Box 4345, Christchurch, New Zealand.

<sup>23</sup>MRC Mitochondrial Biology Unit, Hills Road, Cambridge CB2 0XY, UK.

Received 22 May; accepted 28 August 2014.

- Shen, E.-Z. *et al.* Mitoflash frequency in early adulthood predicts lifespan in *Caenorhabditis elegans*. *Nature* **508**, 128–132 (2014).
- Wang, W. *et al.* Superoxide flashes in single mitochondria. *Cell* **134**, 279–290 (2008).
- Muller, F. L. A critical evaluation of cpYFP as a probe for superoxide. *Free Radic. Biol. Med.* **47**, 1779–1780 (2009).
- Ezeriņa, D., Morgan, B. & Dick, T. P. Imaging dynamic redox processes with genetically encoded probes. *J. Mol. Cell. Cardiol.* **73**, 43–49 (2014).

- Schwarzländer, M., Logan, D. C., Fricker, M. D. & Sweetlove, L. J. The circularly permuted yellow fluorescent protein cpYFP that has been used as a superoxide probe is highly responsive to pH but not superoxide in mitochondria: implications for the existence of superoxide 'flashes'. *Biochem. J.* **437**, 381–387 (2011).
- Santo-Domingo, J., Giacomello, M., Poburko, D., Scorrano, L. & Demareux, N. OPA1 promotes pH flashes that spread between contiguous mitochondria without matrix protein exchange. *EMBO J.* **32**, 1927–1940 (2013).
- Schwarzländer, M. *et al.* Mitochondrial 'flashes': a radical concept repHined. *Trends Cell Biol.* **22**, 503–508 (2012).
- Poburko, D., Santo-Domingo, J. & Demareux, N. Dynamic regulation of the mitochondrial proton gradient during cytosolic calcium elevations. *J. Biol. Chem.* **286**, 11672–11684 (2011).
- Schwarzländer, M. *et al.* Pulsing of membrane potential in individual mitochondria: a stress-induced mechanism to regulate respiratory bioenergetics in *Arabidopsis*. *Plant Cell* **24**, 1188–1201 (2012).
- Azarias, G. & Chatton, J. Y. Selective ion changes during spontaneous mitochondrial transients in intact astrocytes. *PLoS ONE* **6**, e28505 (2011).
- Wei-LaPierre, L. *et al.* Respective contribution of mitochondrial superoxide and pH to mitochondria-targeted circularly permuted yellow fluorescent protein (mt-cpYFP) flash activity. *J. Biol. Chem.* **288**, 10567–10577 (2013).
- Pouvreau, S. Superoxide flashes in mouse skeletal muscle are produced by discrete arrays of active mitochondria operating coherently. *PLoS ONE* **5**, e13035 (2010).
- Zhang, X. *et al.* Superoxide constitutes a major signal of mitochondrial superoxide flash. *Life Sci.* **93**, 178–186 (2013).
- Quatresous, E., Legrand, C. & Pouvreau, S. Mitochondria-targeted cpYFP: pH or superoxide sensor? *J. Gen. Physiol.* **140**, 567–570 (2012).
- Ma, Q. *et al.* Superoxide flashes: early mitochondrial signals for oxidative stress-induced apoptosis. *J. Biol. Chem.* **286**, 27573–27581 (2011).
- Breckwoldt, M. O. *et al.* Multiparametric optical analysis of mitochondrial redox signals during neuronal physiology and pathology in vivo. *Nature Med.* **20**, 555–560 (2014).
- Hou, T., Wang, X., Ma, Q. & Cheng, H. Mitochondrial flashes: new insights into mitochondrial ROS signaling and beyond. *J. Physiol. (Lond.)* **592**, 3703–3713 (2014).

**Author Contributions** M.S., A.J.M., T.P.D. and M.P.M. conceived the project. M.S., S.W., Y.G.E. and V.V.B. performed spectroscopic experiments. S.W. and T.P.D. performed modelling of protein structures. All authors contributed to experimental design, discussed the results, and wrote the manuscript.

**Competing Financial Interests** Declared none.

doi:10.1038/nature13858

## Cheng *et al.* reply

REPLYING TO M. Schwarzländer *et al.* *Nature* **514**, <http://dx.doi.org/10.1038/nature13858> (2014)

In the accompanying Comment<sup>1</sup>, Schwarzländer *et al.* challenged our recent study<sup>2</sup> because they failed to reproduce our previous finding that the fluorescence intensity of purified circularly permuted yellow fluorescent protein (cpYFP) increases in response to oxygen and superoxide anions produced by xanthine (X) plus xanthine oxidase (XO)<sup>3</sup>. Starting from a 'fully reduced' state (incubation with 10 mM dithiothreitol for >3 h) and in the presence of 75 mM HEPES, we demonstrated that cpYFP exhibits a twofold fluorescence increase after oxygenation, and an additional twofold increase after the subsequent addition of X plus XO, which could not be accounted for by solvent (potassium hydroxide)-induced alkalization. Furthermore, the xanthine plus xanthine oxidase-induced increase in cpYFP fluorescence was reversed by Cu/Zn superoxide dismutase (600 U ml<sup>-1</sup>). We also found that the fluorescence intensity of fully reduced cpYFP increased >fourfold after incubation with 1 mM aldrithiol. Notably, recombinant cpYFP purified in the absence of dithiothreitol treatment exhibits a high fluorescence comparable to that of the fully oxidized state, indicating the high susceptibility of cpYFP to oxidation in non-reducing environments<sup>5</sup>. Therefore, ensuring a fully reduced state of cpYFP is essential for the probe to sense superoxide *in vitro*. This property is probably the reason that the probe

functions readily as a reversible superoxide biosensor when targeted to the reduced environment of the mitochondrial matrix. Unfortunately, from the brief description of the methods and limited data provided by Schwarzländer *et al.*<sup>1</sup>, it is not possible to determine whether cpYFP was fully reduced in their experiments, or whether sufficient precautions were taken to prevent oxidation of the probe. Moreover, in our experiments cpYFP was expressed in *Escherichia coli* BL21(DE3)LysS cells, whereas Schwarzländer *et al.*<sup>1</sup> used *E. coli* Origami, a *trxB* (thioredoxin reductase) mutant strain that also lacks glutathione reductase needed to fully limit cysteine oxidation<sup>4</sup>, which could result in an increased oxidative status of their purified cpYFP rendering it non-responsive to superoxide.

Our data from intact cells demonstrate that, in addition to increasing mitoflash frequency, aldrithiol and menadione application also markedly increases basal cpYFP fluorescence intensity within the mitochondrial matrix<sup>2,5</sup>. In addition, nanomolar concentrations of nigericin, a H<sup>+</sup>/K<sup>+</sup> antiporter, stimulates mitoflash activity<sup>6</sup>. These responses of cpYFP *in situ* are unlikely to be attributable to mitochondrial alkalization. We also found that the temporal profile of mitoflash events do not always mirror the change of the mitochondrial membrane potential in



cardiac and skeletal muscle cells<sup>6,7</sup>, and this contradicts the suggestion that mitoflashes simply reflect increased proton pumping in response to membrane potential depolarization<sup>8–10</sup>.

Until now, structural information about how cpYFP senses superoxide remains a mystery. In a unique class of enhanced green fluorescent protein (eGFP)-based calcium sensors, a reversible deprotonization of the chromophore occurs owing to calcium binding to a negatively charged site on the probe<sup>11</sup>. We are investigating whether a similar mechanism might underlie the reversible superoxide-sensing chemistry of cpYFP.

Despite the technical issue raised by Schwarzländer *et al.*<sup>1</sup>, the existence of bursts of superoxide or reactive oxygen species (ROS) production in respiring mitochondria has been confirmed by several independent investigators using different probes. Two pH-insensitive, ROS probes, MitoSOX and 2',7'-dichlorodihydrofluorescein diacetate, have validated ROS increases during cpYFP-reported 'flashes'<sup>6,12,13</sup>. When used individually to avoid possible fluorescence resonance energy transfer (FRET) effects and spectral cross-contamination, these pH-insensitive ROS sensors confirmed flash events of nearly identical frequency and spatiotemporal properties as that observed for cpYFP<sup>13</sup>. Quantification of the respective contributions of superoxide and pH to mitoflash events showed that a predominant superoxide component is coincident with a modest alkalization of the mitochondrial matrix in muscle cells<sup>6</sup>. Similarly, a previous report used MitoSOX to confirm bursts of superoxide during pH alkalization events in primary astrocytes transfected with the fluorescent pH-sensor mitoSypHer<sup>14</sup>. A recent report, which is co-authored by two authors of the accompanying Comment by Schwarzländer *et al.*<sup>1</sup>, used roGFP2 to detect spontaneous, short-lived oxidative bursts that are accompanied by mitochondrial depolarization, transient matrix alkalization, and reversible mitochondrial 'contractions'<sup>15</sup>, all of which we previously documented for cpYFP mitoflashes. Furthermore, in many cell types and tissues<sup>3,5,12,13,16,17</sup> and even in living animals<sup>2</sup>, mitoflash activity is increased by oxidants (including menadione and paraquat) and reduced by antioxidants (including mitoTEMPO and SS31). Nevertheless, given the extreme diversity and plasticity of the mitochondria proteome<sup>18</sup>, the relative contributions of superoxide and pH to cpYFP-reported mitoflash events may vary in a species-, cell-type- and context-dependent manner.

It has become increasingly appreciated that mitoflash activity is a complex phenomenon, comprising multifaceted and intertwined mitochondrial processes including quantal superoxide production, membrane depolarization, membrane permeability transition, NADH depletion, matrix alkalization and swelling that masquerades as mitochondrial contraction<sup>3,6,14,15,17,19</sup>. Ample evidence supports the notion that mitoflash activity serves as a novel and universal "frequency-coded optical readout reflecting free-radical production and energy metabolism at the single-mitochondrion level"<sup>2</sup>. The continuing debate on what drives, controls and contributes to these events does not change the fact that mitoflashes reflect a fundamental physiological phenomenon linked to energy metabolism and stress response, nor does it discount the significance of our finding that mitoflash frequency predicts lifespan at a very early age in *Caenorhabditis elegans*<sup>2</sup>.

The Comment by Schwarzländer *et al.*<sup>1</sup> focuses exclusively on the controversy of cpYFP as a superoxide sensing probe, which was originally demonstrated in several publications by Wang, Dirksen, Sheu and Cheng, and therefore these authors were invited to respond to the Comment. M.-Q. Dong and 11 authors of the original paper<sup>2</sup> were not involved in the research that led to the discovery of cpYFP as a superoxide

probe, so are not listed as authors (M.-Q. Dong was included in this Reply as a corresponding author).

**Heping Cheng<sup>1</sup>, Wang Wang<sup>2</sup>, Xianhua Wang<sup>1</sup>, Shey-Shing Sheu<sup>3</sup>, Robert T. Dirksen<sup>4</sup> & Meng-Qiu Dong<sup>5</sup>**

<sup>1</sup>State Key Laboratory of Biomembrane and Membrane Biotechnology, Beijing Key Laboratory of Cardiometabolic Molecular Medicine, Institute of Molecular Medicine, Peking-Tsinghua Center for Life Sciences, Peking University, Beijing 100871, China.

<sup>2</sup>Mitochondria and Metabolism Center, Department of Anesthesiology and Pain Medicine, University of Washington, Seattle, Washington 98109, USA.

<sup>3</sup>Center for Translational Medicine, Department of Medicine, Sidney Kimmel Medical College, Thomas Jefferson University, Philadelphia, Pennsylvania 19107, USA.

<sup>4</sup>Department of Pharmacology and Physiology, University of Rochester Medical Center, Rochester, New York 14642, USA.

<sup>5</sup>National Institute of Biological Sciences, Beijing, Beijing 102206, China. email: chenghp@pku.edu.cn and dongmengqiu@nibs.ac.cn

- Schwarzländer, M. *et al.* The 'mitoflash' probe cpYFP does not respond to superoxide. *Nature* **514**, <http://dx.doi.org/10.1038/nature13858> (2014).
- Shen, E.-Z. *et al.* Mitoflash frequency in early adulthood predicts lifespan in *Caenorhabditis elegans*. *Nature* **508**, 128–132 (2014).
- Wang, W. *et al.* Superoxide flashes in single mitochondria. *Cell* **134**, 279–290 (2008).
- Derman, A. I., Prinz, W. A., Belin, D. & Beckwith, J. Mutations that allow disulfide bond formation in the cytoplasm of *Escherichia coli*. *Science* **262**, 1744–1747 (1993).
- Huang, Z. *et al.* Response to "A critical evaluation of cpYFP as a probe for superoxide". *Free Radic. Biol. Med.* **51**, 1937–1940 (2011).
- Wei-LaPierre, L. *et al.* Respective contribution of mitochondrial superoxide and pH to mitochondria-targeted circularly permuted yellow fluorescent protein (mt-cpYFP) flash activity. *J. Biol. Chem.* **288**, 10567–10577 (2013).
- Li, K. *et al.* Superoxide flashes reveal novel properties of mitochondrial reactive oxygen species excitability in cardiomyocytes. *Biophys. J.* **102**, 1011–1021 (2012).
- Schwarzländer, M. *et al.* Pulsing of membrane potential in individual mitochondria: a stress-induced mechanism to regulate respiratory bioenergetics in *Arabidopsis*. *Plant Cell* **24**, 1188–1201 (2012).
- Schwarzländer, M. *et al.* Mitochondrial 'flashes': a radical concept repHined. *Trends Cell Biol.* **22**, 503–508 (2012).
- Santo-Domingo, J., Giacomello, M., Poburko, D., Scorrano, L. & Demareux, N. OPA1 promotes pH flashes that spread between contiguous mitochondria without matrix protein exchange. *EMBO J.* **32**, 1927–1940 (2013).
- Tang, S. *et al.* Design and application of a class of sensors to monitor Ca<sup>2+</sup> dynamics in high Ca<sup>2+</sup> concentration cellular compartments. *Proc. Natl Acad. Sci. USA* **108**, 16265–16270 (2011).
- Pouvreau, S. Superoxide flashes in mouse skeletal muscle are produced by discrete arrays of active mitochondria operating coherently. *PLoS ONE* **5**, e13035 (2010).
- Zhang, X. *et al.* Superoxide constitutes a major signal of mitochondrial superoxide flash. *Life Sci.* **93**, 178–186 (2013).
- Azarias, G. & Chatton, J. Y. Selective ion changes during spontaneous mitochondrial transients in intact astrocytes. *PLoS ONE* **6**, e28505 (2011).
- Breckwoldt, M. O. *et al.* Multiparametric optical analysis of mitochondrial redox signals during neuronal physiology and pathology *in vivo*. *Nature Med.* **20**, 555–560 (2014).
- Hou, T. *et al.* Synergistic triggering of superoxide flashes by mitochondrial Ca<sup>2+</sup> uniport and basal reactive oxygen species elevation. *J. Biol. Chem.* **288**, 4602–4612 (2013).
- Ma, Q. *et al.* Superoxide flashes: early mitochondrial signals for oxidative stress-induced apoptosis. *J. Biol. Chem.* **286**, 27573–27581 (2011).
- Pagliarini, D. J. *et al.* A mitochondrial protein compendium elucidates complex I disease biology. *Cell* **134**, 112–123 (2008).
- Hou, T., Wang, X., Ma, Q. & Cheng, H. Mitochondrial flashes: new insights into mitochondrial ROS signaling and beyond. *J. Physiol. (Lond.)* **592**, 3703–3713 (2014).

doi:10.1038/nature13859

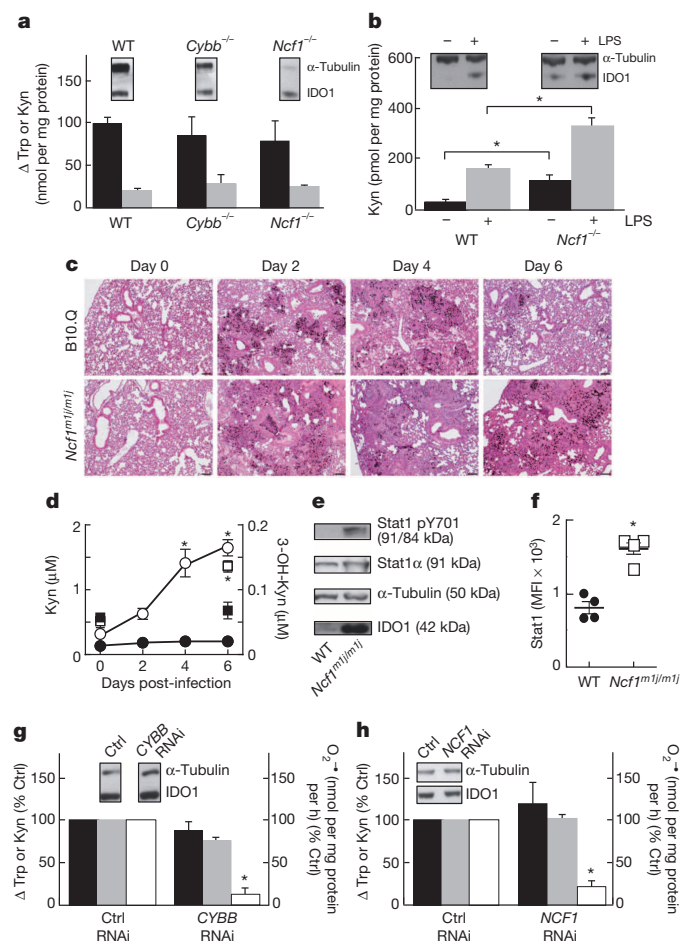
# Tryptophan catabolism is unaffected in chronic granulomatous disease

ARISING FROM L. Romani *et al.* *Nature* **451**, 211–215 (2008); doi:10.1038/nature06471

Chronic granulomatous disease (CGD) is an inherited disorder of phagocyte function, caused by a genetic defect in NADPH oxidase (NOX2), leading to an impaired ability of leukocytes to produce superoxide ( $O_2^{\cdot-}$ ); CGD subjects are susceptible to chronic infections and hyperinflammation, although the mechanisms remain unclear. Romani *et al.*<sup>2</sup> reported an aberrant inflammatory response to pulmonary aspergillosis as well as sterile *Aspergillus fumigatus* to be mediated by a defective tryptophan catabolism to kynurenine caused by lack of  $O_2^{\cdot-}$  in CGD mice. Kynurenine is formed by indoleamine 2,3-dioxygenase-1 (IDO1) in a reaction originally reported to depend on  $O_2^{\cdot-}$  (ref. 3). Here we show that NOX2 deficiency does not attenuate IDO1-mediated tryptophan catabolism in human phagocytes and CGD mice with granulomas arising from an inflammatory response to *Aspergillus*. There is a Reply to this Brief Communications Arising by Romani, L. & Puccetti, P. *Nature* **514**, http://dx.doi.org/10.1038/nature13845 (2014).

Romani *et al.*<sup>2</sup> concluded that IDO-mediated tryptophan catabolism is blocked in CGD based on studies performed in *p47<sup>phox</sup>*<sup>-/-</sup> mice. They reported increased kynurenine in granuloma-containing lungs of wild-type but not *p47<sup>phox</sup>*<sup>-/-</sup> mice, and that interferon- $\gamma$  (IFN- $\gamma$ ) stimulates IDO activity in lung phagocytes from wild-type but not *p47<sup>phox</sup>*<sup>-/-</sup> mice, despite the presence of IDO protein in both mouse strains.

However, recent studies have shown that cytochrome  $b_5$  rather than  $O_2^{\cdot-}$  activates cellular IDO<sup>4</sup>. Therefore, we re-examined tryptophan catabolism to kynurenine in several models of CGD, using *Ncf1*<sup>-/-</sup> (lacking



**Figure 1 | Phagocyte NADPH oxidase activity is not required for IDO1 activity in inflammation.** **a**, Lung PMN (purity > 85% by differential count) were isolated from wild-type (WT), *Ncf1*<sup>-/-</sup> or *Cybb*<sup>-/-</sup> mice 24 h after intra-peritoneal injection of 7.5 mg per kg lipopolysaccharide 0111:B4 (LPS). PMN were then incubated for 48 h in RPMI medium containing 10% fetal bovine serum in the presence of 200 U ml<sup>-1</sup> mouse recombinant IFN- $\gamma$  and 400  $\mu$ M L-N<sup>G</sup>-monomethyl-arginine (nitric oxide synthase inhibitor). Consumption of tryptophan ( $\Delta$ Trp, black) and accumulation of kynurenine (Kyn, grey) in the medium were then determined by high-performance liquid chromatography (HPLC) (mean  $\pm$  s.e.m., *n* = 3); insets show representative  $\alpha$ -tubulin (50 kDa) and IDO1 (42 kDa) proteins. PMN from CGD mice failed to generate  $O_2^{\cdot-}$  upon stimulation with IFN- $\gamma$  ( $72 \pm 9$  versus  $0 \pm 0$  nmol  $O_2^{\cdot-}$  per h per mg protein for WT versus *Ncf1*<sup>-/-</sup> or *Cybb*<sup>-/-</sup>), mean  $\pm$  s.e.m., *n* = 3 (data not shown).

**b**, Lungs from WT and *Ncf1*<sup>-/-</sup> mice before (-) and 24 h after intra-peritoneal administration of 7.5 mg per kg LPS (+) were homogenized in 1 ml of 20 mM phosphate buffer pH 7.2 containing 140 mM KCl and 2 $\times$  complete protease inhibitor cocktail (Roche), and Kyn determined by HPLC; insets show representative  $\alpha$ -tubulin (50 kDa) and IDO1 (42 kDa) proteins. Data are mean  $\pm$  s.e.m. of 3 mice in each group; \**P* < 0.05 (Mann-Whitney rank sum test). **c**, Representative lung sections (5- $\mu$ m thick) prepared from B10.Q and *Ncf1*<sup>mlj/mlj</sup> mice before (day 0) and 2, 4 and 6 days after intra-nasal instillation of 5  $\mu$ g of sterile hyphal cell wall from *A. fumigatus*, and stained with haematoxylin and eosin. Scale bar, 100  $\mu$ m. **d**, Lung Kyn (circles) and 3-hydroxykynurenine (3-OH-Kyn, squares) in B10.Q (WT, black symbols) and *Ncf1*<sup>mlj/mlj</sup> mice (white symbols) before (day 0) and 2, 4 and 6 days after instillation of sterile hyphal cell wall of *A. fumigatus*. Lungs were homogenized, proteins precipitated with 4% trichloroacetic acid, the mixture centrifuged and the resulting supernatant neutralised (1 M sodium phosphate, pH 7.4) and subjected to liquid chromatography with tandem mass spectrometry (LC/MS/MS) using *m/z* 209 $\rightarrow$ 146 and 225 $\rightarrow$ 208 transitions for Kyn and 3-OH-Kyn, respectively.

Kyn and 3-OH-Kyn were separated on a 150  $\times$  4.6 mm Luna C18 (2), 5  $\mu$ m column (Phenomenex) using a gradient generated by mobile phase A (0.1% formic acid) and B (0.1% formic acid in 100% acetonitrile). The results show mean  $\pm$  s.e.m., with *n* = 7–8 for Kyn and *n* = 4 for 3-OH-Kyn for each time point and genotype. \**P* < 0.05 indicates significant difference between *Ncf1*<sup>mlj/mlj</sup> and B10.Q using two-way ANOVA followed by Sidak's multiple comparison test. **e**, Representative Western blots of lung homogenates from B10.Q (WT) and *Ncf1*<sup>mlj/mlj</sup> mice 6 days after instillation of sterile hyphal cell wall from *A. fumigatus*, showing phosphorylated Stat1 (Stat1 pY701; 91 and 84 kDa), Stat1 $\alpha$  (91 kDa),  $\alpha$ -tubulin (50 kDa) and IDO1 (42 kDa) proteins.

**f**, Cells from bronchoalveolar lavage fluid of B10.Q (WT) and *Ncf1*<sup>mlj/mlj</sup> mice 4 days after instillation of *A. fumigatus* were stained by cell surface markers, fixed and incubated with anti-Stat1 antibody (BD Biosciences). Expression of Stat1 in CD11b<sup>+</sup>Ly6g<sup>+</sup> PMN was assessed by flow cytometry and is shown as median fluorescence intensity (MFI). Data are shown for individual animals as well as mean  $\pm$  s.e.m. \**P* < 0.05 (Mann-Whitney rank sum test). **g**, **h**, THP-1 cells ( $10^6$  per well) were treated for 24 h with the respective control, *CYBB*-specific siRNA (5'-CGGAGGUUUACUUUGAAGUCUUU-3', Invitrogen) (**g**) or *NCF1*-specific siRNA (sc-29422; Santa Cruz Biotechnologies) (**h**), followed by 48 h incubation in the presence of 400 U ml<sup>-1</sup> recombinant human IFN- $\gamma$ .

Trp lost from ( $\Delta$ Trp, black) and Kyn accumulated in the medium (grey) was then determined by HPLC. Generation of  $O_2^{\cdot-}$  (white) was determined by cytochrome *c* reduction after treatment of the cells with 200 ng ml<sup>-1</sup> phorbol-12-myristate-13-acetate for 1 h. The 100% control-values for **g** and **h** were  $15 \pm 3.4$  and  $51 \pm 9$  nmol  $O_2^{\cdot-}$  per h per mg protein, respectively (mean  $\pm$  s.e.m., *n* = 3); insets show representative  $\alpha$ -tubulin (50 kDa) and IDO1 (42 kDa) proteins. \**P* < 0.05 (Mann-Whitney rank sum test).



the Ncf1 protein, also known as p47<sup>phox</sup>, Cybb<sup>-/-</sup> (lacking the catalytic subunit of NOX2, also known as gp91<sup>phox</sup>)<sup>5</sup> and Ncf1<sup>m1j/m1j</sup> mice (with a single mutation in the Ncf1 gene, resulting in a defective NCF1 protein leading to a lack of NOX2 activity)<sup>6,7</sup>.

Polymorphonuclear leukocytes (PMN) from lungs of endotoxin-treated Ncf1<sup>-/-</sup> and Cybb<sup>-/-</sup> mice failed to generate O<sub>2</sub><sup>-•</sup> upon activation, yet these cells, like wild-type cells, converted tryptophan to kynurenine (Fig. 1a). Moreover, endotoxin treatment increased pulmonary kynurenine similarly in wild-type and Ncf1<sup>-/-</sup> mice (Fig. 1b). To examine tryptophan catabolism in situations of hyper-inflammation, we administered sterile *Aspergillus fumigatus* to Ncf1<sup>m1j/m1j</sup> and their wild-type control mice (B10.Q), differing at only a single Ncf1 mutation. As expected, this caused pulmonary granulomas in B10.Q and Ncf1<sup>m1j/m1j</sup> mice, and these granulomas resolved only in control animals (Fig. 1c). Strikingly, lung IDO1 protein, kynurenine as well as 3-hydroxykynurenine were higher with defective NCF1 (Fig. 1d, e). This was associated with an increase in pulmonary and bronchoalveolar lavage fluid PMN Stat1 protein (Fig. 1e, f), and lung phosphorylated Stat1 (Fig. 1e). Phosphorylation of Stat1 is a major NOX2 downstream pathway<sup>8</sup> that mediates IFN-γ-dependent IDO1 expression<sup>9</sup>.

The above studies imply that O<sub>2</sub><sup>-•</sup> is not required for *in vivo* IDO activity. Consistent with this, blood PMN isolated from CGD patients with a mutation in the NCF1 or CYBB gene were unable to generate O<sub>2</sub><sup>-•</sup>, yet effectively degraded tryptophan to kynurenine (data not shown), as also reported recently by others<sup>10,11</sup>. Similarly, knockdown of NCF1 or CYBB protein in human monocytic THP-1 cells did not decrease tryptophan catabolism to kynurenine, although it blunted O<sub>2</sub><sup>-•</sup> formation by ~80% (Fig. 1g, h).

Our observation of increased, as opposed to decreased<sup>2</sup>, IDO activity and 3-hydroxykynurenine (an indicator of kynurenine metabolism) in infected CGD mice is consistent with studies reporting elevated plasma and urinary kynurenine and 3-hydroxykynurenine in CGD patients<sup>10,12</sup>. Also, gene therapy with CYBB in a CGD patient resulted in clearance of *Aspergillus* infection that was associated with restoration of 30% of normal NOX activity without increase of IDO activity, as assessed by plasma kynurenine<sup>13</sup>. The discrepancy between our cellular studies (Fig. 1a) and those of Romani *et al.*<sup>2</sup> may be explained in part by nitric oxide inhibiting IDO activity<sup>14</sup>, as we observed IDO activity only when nitric oxide synthases were blocked. Moreover, the Ncf1<sup>m1j/m1j</sup> mice used here<sup>7</sup> better reflect human CGD in which NOX proteins are expressed, albeit as non-functional mutants, rather than being absent and with linked chromosomal fragment/s from the original 129 derived embryonic stem cell, which could also vary in different backcrossed strains, as in the case of the Ncf1<sup>-/-</sup> mice used<sup>15</sup>.

We conclude that IDO1-mediated tryptophan catabolism to kynurenine does not require phagocyte NADPH oxidase derived O<sub>2</sub><sup>-•</sup>, nor is it defective in human or mouse CGD, even under conditions where hyper-inflammation exists. Therefore, blockade of this pathway is unlikely to explain the acute pulmonary inflammatory response observed by Romani *et al.*<sup>2</sup>.

## Methods

Lungs and lung PMN were isolated from mice and cells cultured as described previously<sup>2,14</sup>. Sterile *Aspergillus fumigatus* hyphal cell wall was administered intranasally to B10.Q and Ncf1<sup>m1j/m1j</sup> mice<sup>5</sup>. Human blood PMN were isolated from control or CGD patients by a Percoll gradient<sup>16</sup>. Knockdown of NCF1 or CYBB in THP-1 cells was achieved by short interfering RNA (siRNA) transfection. Tryptophan, kynurenine and 3-hydroxykynurenine in the medium of IFN-γ-treated cells and in lung homogenates were quantified by high-performance liquid chromatography<sup>4</sup> and liquid chromatography–tandem mass spectrometry (LC/MS/MS), respectively. Cellular NOX activity was determined by cytochrome *c* reduction following stimulation with phorbol-12-myristate-13-acetate.

All experiments involving animal and human subjects were performed with approval from the institutional ethics committees. We thank the CGD donors for

their co-operation and donation of blood, and J. Ziegler for his help. We also thank E. Lönnblom, E. Chrysanthou and K. Selva Nandakumar for their help with the *Aspergillus* experiments. This work was supported by grants from the National Health and Medical Research Council of Australia (1003484 and 1037879 to RS and 455395 to R.S. and B.H.C.) and the European Commission Directorate-General for Research & Innovation (HEALTH-F2-2012-278611).

**Ghassan J. Maghzal<sup>1,2,3</sup>, Susann Winter<sup>4</sup>, Bettina Wurzer<sup>3</sup>, Beng H. Chong<sup>5</sup>, Rikard Holmdahl<sup>4</sup> & Roland Stocker<sup>1,2,3</sup>**

<sup>1</sup>Vascular Biology Division, Victor Chang Cardiac Research Institute, Darlinghurst, New South Wales 2010, Australia.

email: r.stocker@victorchang.edu.au

<sup>2</sup>School of Medicine, University of New South Wales, New South Wales 2052, Australia.

<sup>3</sup>School of Medical Sciences and Bosch Institute, The University of Sydney, Camperdown, New South Wales 2050, Australia.

<sup>4</sup>Department of Medical Biochemistry and Biophysics, Karolinska Institutet, 171 77 Stockholm, Sweden.

<sup>5</sup>Centre for Vascular Research, University of New South Wales, New South Wales 2052, Australia.

**Received 26 June; accepted 4 August 2014.**

- Segal, B. H., Leto, T. L., Gallin, J. I., Malech, H. L. & Holland, S. M. Genetic, biochemical, and clinical features of chronic granulomatous disease. *Medicine* **79**, 170–200 (2000).
- Romani, L. *et al.* Defective tryptophan catabolism underlies inflammation in mouse chronic granulomatous disease. *Nature* **451**, 211–215 (2008).
- Hirata, F. & Hayaishi, O. Studies on indoleamine 2,3-dioxygenase. I. Superoxide anion as substrate. *J. Biol. Chem.* **250**, 5960–5966 (1975).
- Maghzal, G. J., Thomas, S. R., Hunt, N. H. & Stocker, R. Cytochrome b<sub>5</sub>, not superoxide anion radical, is a major reductant of indoleamine 2,3-dioxygenase in human cells. *J. Biol. Chem.* **283**, 12014–12025 (2008).
- Morgenstern, D. E., Gifford, M. A., Li, L. L., Doerschuk, C. M. & Dinan, M. C. Absence of respiratory burst in X-linked chronic granulomatous disease mice leads to abnormalities in both host defense and inflammatory response to *Aspergillus fumigatus*. *J. Exp. Med.* **185**, 207–218 (1997).
- Hultqvist, M. *et al.* Enhanced autoimmunity, arthritis, and encephalomyelitis in mice with a reduced oxidative burst due to a mutation in the Ncf1 gene. *Proc. Natl Acad. Sci. USA* **101**, 12646–12651 (2004).
- Sareila, O., Jaakkola, N., Olofsson, P., Kelkka, T. & Holmdahl, R. Identification of a region in p47phox/NCF1 crucial for phagocytic NADPH oxidase (NOX2) activation. *J. Leukoc. Biol.* **93**, 427–435 (2013).
- Kelkka, T. *et al.* Reactive oxygen species deficiency induces autoimmunity with type 1 interferon signature. *Antioxid. Redox Signal.* <http://dx.doi.org/10.1089/ars.2013.5828> (29 July 2014).
- Thomas, S. R. & Stocker, R. Redox reactions related to indoleamine 2,3-dioxygenase and tryptophan metabolism along the kynurenine pathway. *Redox Rep.* **4**, 199–220 (1999).
- De Ravin, S. S. *et al.* Tryptophan/kynurenine metabolism in human leukocytes is independent of superoxide and is fully maintained in chronic granulomatous disease. *Blood* **116**, 1755–1760 (2010).
- Jürgens, B., Fuchs, D., Reichenbach, J. & Heitger, A. Intact indoleamine 2,3-dioxygenase activity in human chronic granulomatous disease. *Clin. Immunol.* **137**, 1–4 (2010).
- Heeley, A. F., Heeley, M. E., Hardy, J. & Soothill, J. F. A disorder of tryptophan metabolism in chronic granulomatous disease. *Arch. Dis. Child.* **45**, 485–490 (1970).
- Hakim, A. *et al.* Response: protecting against *Aspergillus* infection in CGD. *Blood* **114**, 3498 (2009).
- Thomas, S. R., Mohr, D. & Stocker, R. Nitric oxide inhibits indoleamine 2,3-dioxygenase activity in interferon-primed mononuclear phagocytes. *J. Biol. Chem.* **269**, 14457–14464 (1994).
- Jackson, S. H., Gallin, J. I. & Holland, S. M. The p47<sup>phox</sup> mouse knock-out model of chronic granulomatous disease. *J. Exp. Med.* **182**, 751–758 (1995).
- Stocker, R., Winterhalter, K. H. & Richter, C. Increased fluorescence polarization of 1,6-diphenyl-1,3,5-hexatriene in the phorbol myristate acetate-stimulated plasma membrane of human neutrophils. *FEBS Lett.* **144**, 199–203 (1982).

**Author Contributions** G.J.M. designed and carried out most experiments. S.W. and R.H. designed and carried out studies involving *A. fumigatus*. B.W. carried out initial cellular studies and B.H.C. was responsible for studies involving CGD patients. R.S. conceived the study and wrote the manuscript with G.J.M. All authors read and contributed to the final version of the manuscript.

**Competing Financial Interests** Declared none.

doi:10.1038/nature13844

# Romani & Puccetti reply

REPLYING TO G. J. Maghzal *et al.* *Nature* **514**, <http://dx.doi.org/10.1038/nature13844> (2014)

After our initial observation of defective tryptophan catabolism in experimental chronic granulomatous disease (CGD)<sup>1</sup>, several laboratories have been testing the indoleamine 2,3-dioxygenase (IDO1) competence of cells from CGD patients. In most instances, they found no impairment in IDO1 competence in terms of tryptophan catabolic activity *in vitro* by polymorphonuclear leukocytes and monocyte-derived dendritic cells<sup>2,3</sup>, leading to the conclusion that there is no obvious defect in the production of kynurenine (the first by-product of tryptophan degradation)—hence in the IDO1-dependent mechanism of tolerogenesis as a whole in human CGD. In the accompanying Comment<sup>4</sup>, Maghzal *et al.* report that tryptophan catabolism is unaffected in chronic granulomatous disease, again by measurements of kynurenine production.

However, a number of studies have now been providing evidence that the assay is not sufficiently informative as to the local versus systemic functioning of the IDO1 mechanism, particularly on considering the pleiotropic effects of IDO1 *in vivo*<sup>5,6</sup>. While we stand by our original observations in CGD mice with lethal pulmonary aspergillosis<sup>1</sup>, it should be noted that, in all of the experimental models tested so far, lack of IDO1 competence does not result, *per se*, in spontaneous inflammatory pathology. Yet, the functional defect becomes obvious when mice not competent for IDO1 function are challenged with an inflammatory *noxa* recognized more usually by Toll-like receptors<sup>7</sup>. Additionally, a number of potential factors have now been identified that further substantiate the concept that a global defect in IDO1 functioning underlies the severe chronic inflammation in CGD, among which are local accumulation of peroxynitrites<sup>8</sup> (which compromise IFN- $\gamma$  signalling necessary for IDO1 induction<sup>9</sup>) and IL-6 (ref. 8) (which promotes IDO1 proteasomal degradation<sup>10</sup>), lack of IDO1-dependent neutrophil apoptosis<sup>11</sup>, loss of IDO1-driven non-canonical NF- $\kappa$ B activation (otherwise resulting in downregulation of proinflammatory cytokines, and upregulation of tolerogenic TGF- $\beta$ ), and probably defective IDO1 signalling activity<sup>12</sup>.

Thus the problem is not whether cells from CGD patients and mice equally display defective tryptophan catabolism *in vitro*. Rather, when contextualized to the current knowledge, that is a matter of appreciation that the IDO1 mechanism, and the multiple downstream regulatory responses over which it presides—including control of the over-reactive responses to TLR signalling—are globally compromised. The situation—that is, elevated rather than suppressed circulating kynurenine and/or *in vitro* production in CGD patients<sup>3</sup>—may be similar to that of septic patients who display high levels of circulating kynurenine in the face of defective overall IDO1 functioning<sup>13</sup>.

Substantial differences might occur between human and experimental CGD, and p47<sup>phox</sup>-deficient mice with infection-related acute inflammatory lung injury may well be an extreme condition. Such as they are, those mice provide a sound proof-of-principle in their being a prototypic

model of a specific condition, exemplifying how in experimental CGD the IDO1 mechanism of disease tolerance<sup>13</sup> is severely compromised at sites where it would mostly be beneficial to control local inflammatory reactions. The very nature of granulation formation could be but one of the multiple phenotypic manifestations of defective IDO1 functioning in infection-related pathology<sup>14</sup>. This Reply has been written on behalf of the entire original author list<sup>1</sup>, most of the original authors are no longer working with us or on the specific subject.

**L. Romani<sup>1</sup> & P. Puccetti<sup>2</sup>**

<sup>1</sup>Department of Experimental Medicine, University of Perugia, 06132 Perugia, Italy.

<sup>2</sup>Department of Experimental Medicine, University of Perugia, 06132 Perugia, Italy.

email: plopcc@tin.it

1. Romani, L. *et al.* Defective tryptophan catabolism underlies inflammation in mouse chronic granulomatous disease. *Nature* **451**, 211–215 (2008).
2. Jürgens, B., Fuchs, D., Reichenbach, J. & Heitger, A. Intact indoleamine 2,3-dioxygenase activity in human chronic granulomatous disease. *Clin. Immunol.* **137**, 1–4 (2010).
3. De Ravin, S. S. *et al.* Tryptophan/kynurenine metabolism in human leukocytes is independent of superoxide and is fully maintained in chronic granulomatous disease. *Blood* **116**, 1755–1760 (2010).
4. Maghzal, G. J. *et al.* Tryptophan catabolism is unaffected in chronic granulomatous disease. *Nature* **514**, <http://dx.doi.org/10.1038/nature13844> (2014).
5. Fallarino, F. *et al.* The combined effects of tryptophan starvation and tryptophan catabolites down-regulate T cell receptor  $\zeta$ -chain and induce a regulatory phenotype in naive T cells. *J. Immunol.* **176**, 6752–6761 (2006).
6. Fallarino, F., Grohmann, U. & Puccetti, P. Indoleamine 2,3-dioxygenase: from catalyst to signaling function. *Eur. J. Immunol.* **42**, 1932–1937 (2012).
7. Volpi, C., Mondanelli, G., Puccetti, P. & Grohmann, U. TLRs and tryptophan metabolism at the crossroad of immunoregulatory pathways. *Immunometabolism* **1**, 28–50 (2014).
8. Rodrigues-Sousa, T. *et al.* Deficient production of reactive oxygen species leads to severe chronic DSS-induced colitis in Ncf1/p47<sup>phox</sup>-mutant mice. *PLoS ONE* **9**, e97532 (2014).
9. Grohmann, U. *et al.* A defect in tryptophan catabolism impairs tolerance in nonobese diabetic mice. *J. Exp. Med.* **198**, 153–160 (2003).
10. Orabona, C. *et al.* SOCS3 drives proteasomal degradation of indoleamine 2,3-dioxygenase (IDO) and antagonizes IDO-dependent tolerogenesis. *Proc. Natl Acad. Sci. USA* **105**, 20828–20833 (2008).
11. Röhm, M. *et al.* NADPH oxidase promotes neutrophil extracellular trap formation in pulmonary aspergillosis. *Infect. Immun.* **82**, 1766–1777 (2014).
12. Pallotta, M. T. *et al.* Indoleamine 2,3-dioxygenase is a signaling protein in long-term tolerance by dendritic cells. *Nature Immunol.* **12**, 870–878 (2011).
13. Bessede, A. *et al.* Aryl hydrocarbon receptor control of a disease tolerance defence pathway. *Nature* **511**, 184–190 (2014).
14. Zelante, T., Fallarino, F., Biston, F., Puccetti, P. & Romani, L. Indoleamine 2,3-dioxygenase in infection: the paradox of an evasive strategy that benefits the host. *Microbes Infect.* **11**, 133–141 (2009).

doi:10.1038/nature13845



## A crack in the natural-gas bridge

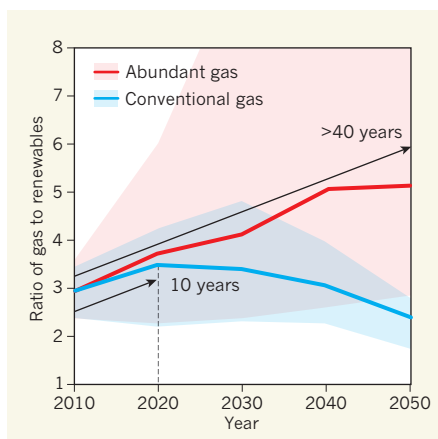
Integrated assessment models show that, without new climate policies, abundant supplies of natural gas will have little impact on greenhouse-gas emissions and climate change. [SEE LETTER P.482](#)

STEVEN J. DAVIS & CHRISTINE SHEARER

Burning fossil fuels such as coal, gas and oil produces more than 80% of the world's energy and more than 90% of global carbon dioxide emissions. Slowing and ultimately stopping climate change depends on decarbonization — the transformation of the global energy system into one that does not dump CO<sub>2</sub> into the atmosphere. Because gas-fired power plants emit roughly half as much CO<sub>2</sub> per unit of energy produced as coal-fired plants, the greatly expanded gas supplies promised by new hydraulic fracturing (fracking) methods have been celebrated as a means of cutting emissions<sup>1</sup>. Progressive substitution of gas for coal and oil can thus decarbonize the energy sector<sup>2</sup> and serve as a 'bridge' to a more distant future when carbon-free, renewable-energy technologies are more affordable and reliable than they are now<sup>3</sup>. In this issue, McJeon *et al.*<sup>4</sup> (page 482) uncover a serious crack in the gas bridge: in the absence of new climate policies, increased supplies of natural gas may have little effect on CO<sub>2</sub> emissions and could actually delay decarbonization of the global energy system.

McJeon and colleagues' findings reveal two effects. First, abundant gas makes energy cheaper, thereby encouraging higher energy consumption and discouraging investment in energy efficiency. Second, natural gas competes for market share not only with coal, but also with very-low-carbon energy sources such as renewables and nuclear.

Previous studies have questioned the climate benefits of natural gas relative to coal owing to the potential for the gas (mostly methane, a greenhouse gas) to leak into the atmosphere during its extraction, processing and transport<sup>5</sup>. More recently, researchers have begun to consider the effects of abundant natural gas on CO<sub>2</sub> emissions in the broader context of the energy marketplace<sup>6–9</sup>. McJeon and co-workers' paper is the first peer-reviewed study to do so on a global scale. It uses five independent energy-economic models to simulate the effects of gas supplies on the global energy system and on emissions of CO<sub>2</sub>, methane, nitrous oxide and aerosols such as sulphur dioxide and black carbon. Their study compares a 'conventional' gas supply with an 'abundant' case in which



**Figure 1 | Relative growth of gas and renewable electricity.** The ratio of natural gas to renewables used to generate electricity is sensitive to how much inexpensive gas is available. The red and blue lines show the median of this ratio across five energy-economic models used by McJeon *et al.*<sup>4</sup> for scenarios of abundant and conventional gas supplies, respectively, whereas the shaded areas show the full range spanned by the individual models. For cases in which less gas is available (that is, in the conventional scenario), renewables as an electricity source begin to grow faster than gas 10 years into the 40-year modelling period. But when gas is abundant, its use grows faster than that of renewables throughout the period modelled and probably beyond it.

natural-gas prices are halved, and evaluates the net influence of emissions on the climate system in the two scenarios.

In all five models used by the authors, CO<sub>2</sub> emissions and their effect on climate (climate forcing) scarcely differed between the conventional and abundant scenarios. At most, abundant gas reduced cumulative CO<sub>2</sub> emissions between 2010 and 2050 by 2%, and reduced climate forcing over the same period by 0.3%. In several of the models, emissions and forcing actually increased under the abundant-gas scenario. But the exact numbers, although revealing, are less important than the overall insight: whether the goal is avoiding CO<sub>2</sub> emissions or hastening the transition to an emissions-free energy system, a global gas boom is not a replacement for energy and climate policies.

Indeed, by replottting some of McJeon and

colleagues' results, it is possible to observe the extent to which the availability of abundant gas delays the transition to low-carbon, renewable energy sources such as solar and wind. Figure 1 shows the ratio of the amount of gas to renewables used to generate electricity in the authors' models between 2010 and 2050. In the race between fossil fuels and low-carbon energy, the lines in the figure (which reflect the median of all five models) indicate which energy source is gaining ground. In the abundant-gas scenario, the ratio never decreases: gas-fired power pulls further and further ahead of renewable power throughout the 40-year period. But in the conventional-gas scenario, the ratio begins to decrease from 2020: renewables start catching up.

McJeon and co-workers' study assumes that there will be no policies intended to reduce greenhouse-gas emissions or to support low-carbon energy other than those already in place. Future work must carefully assess the effectiveness of various policies in reducing greenhouse-gas emissions and decarbonizing the global energy system. Similarly, the authors' results are probably sensitive to assumptions about the cost of low-carbon energy technologies over time, and systematic analyses of such sensitivity could inform energy funding and policies. Finally, further studies may be needed to evaluate the extent to which natural gas could be used strategically to complement and support variable renewable-energy technologies by providing flexible back-up power that can ramp up quickly<sup>10</sup>. Such applications could have very different implications for decarbonization and cumulative CO<sub>2</sub> emissions. Rather than simply building vast fleets of gas-fired power plants that lock in another generation of "committed emissions"<sup>11</sup>, if we get the technologies and the policies right, natural gas might help us to cut emissions by working with renewable energy sources, rather than against them.

The integrated analysis of McJeon *et al.* makes it clear that emissions per unit of energy is a poor measure of prospective energy sources. Differences in emissions between energy sources, considered in isolation, may be irrelevant given the complex feedbacks of the energy markets. Specifically, the authors' study is the most robust evidence yet that expanding

supplies of natural gas will not help us to avoid climate change and manage the transition to renewable energy sources in the absence of an effective climate policy. ■

Steven J. Davis and Christine Shearer are in the Department of Earth System Science, University of California, Irvine, Irvine, California 92697, USA.  
e-mails: [sjdavis@uci.edu](mailto:sjdavis@uci.edu); [cesheare@uci.edu](mailto:cesheare@uci.edu)

## HIV

# A stamp on the envelope

A high-resolution crystal structure of the HIV-1 Env trimer proteins, in their form before they fuse with target cells, will aid the design of vaccines that elicit protective immune responses to this protein complex. [SEE ARTICLE P.455](#)

ROGIER W. SANDERS & JOHN P. MOORE

The surface of the HIV-1 virus is studded with envelope glycoprotein (Env) spikes. The virus uses these trimeric complexes, which each contain three gp120 and three gp41 subunits, to fuse with cells and initiate infection. Sixteen years ago, Kwong and colleagues described the crystal structure of the core of the gp120 subunit<sup>1</sup>, but the lability of the complex in its pre-fusion form meant that the trimer structure was not determined until last year, when an engineered, stabilized and soluble version was used to produce highly concordant structures by X-ray crystallography<sup>2</sup> and cryo-electron

1. [www.bbc.com/news/uk-politics-25705550](http://www.bbc.com/news/uk-politics-25705550)
2. Ausubel, J. H., Grubler, A. & Nakicenovic, N. *Clim. Change* **12**, 245–263 (1988).
3. Podesta, J. D. & Wirth, T. E. *Natural Gas: A Bridge Fuel for the 21st Century* (Center for American Progress, 2009); <http://cdn.americanprogress.org/wp-content/uploads/issues/2009/08/pdf/naturalgasmemo.pdf>
4. McJeon, H. et al. *Nature* **514**, 482–485 (2014).
5. Brandt, A. R. et al. *Science* **343**, 733–735 (2014).
6. Huntington, H. *EMF26: Changing the Game? Emissions and Market Implications of New Natural Gas Supplies* (Energy Modeling Forum, 2013).

7. Newell, R. G. & Raimi, D. *Environ. Sci. Technol.* **48**, 8360–8368 (2014).
8. Shearer, C., Bistline, J., Inman, M. & Davis, S. J. *Environ. Res. Lett.* **9**, 094008 (2014).
9. International Energy Agency. *Golden Rules for a Golden Age of Gas* (IEA, 2012).
10. Channell, J., Lam, T. & Pourreza, S. *Shale & Renewables: A Symbiotic Relationship* (Citi Res., 2012).
11. Davis, S. J. & Socolow, R. H. *Environ. Res. Lett.* **9**, 084018 (2014).

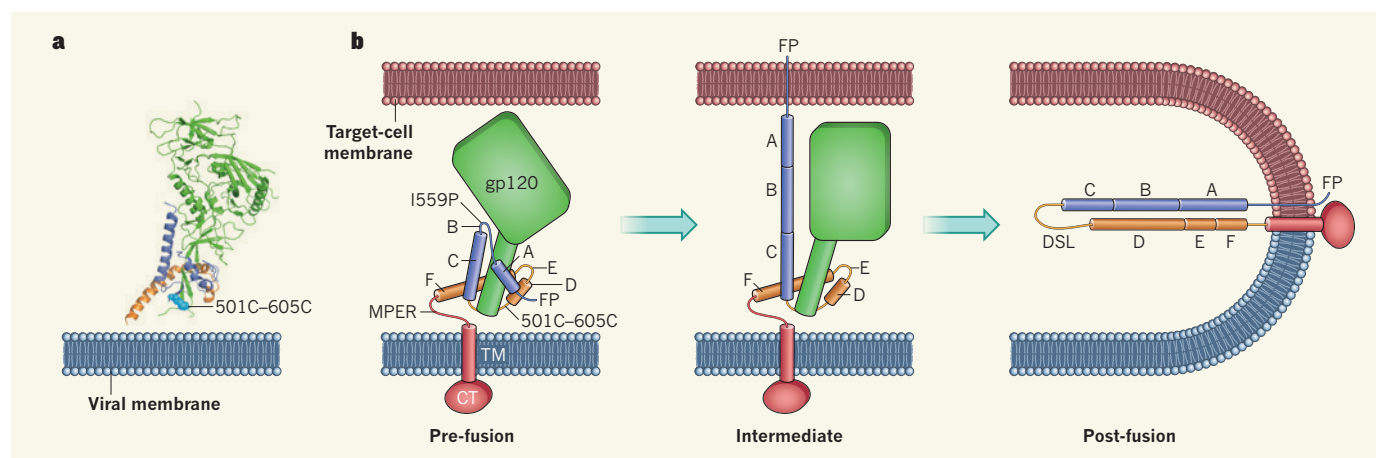
This article was published online on 15 October 2014.

microscopy<sup>3</sup>. Now, on page 455 of this issue, Kwong and colleagues (Pancera et al.)<sup>4</sup> report a crystal structure of the same Env trimer at higher resolution, providing a better picture, particularly of the gp41 subunits. Together, these structures<sup>2–4</sup> help us to understand how the Env trimer functions, how antibodies recognize it (or not), and how vaccines exploiting this protein can be better designed.

HIV-1 fusion occurs when the gp120 components of Env trimers interact first with CD4 receptors on a cell's surface and then with a co-receptor (CCR5 or CXCR4). The sequential receptor engagements drive the concerted disentanglement of the intimate, but fragile, embrace between gp120 and gp41. The

ectodomain of each gp41 subunit (the region that extends out from the viral membrane) contains six segments (A–F) that form two heptad-repeat regions (HR1 and HR2). These segments eventually become two long helices in the post-fusion structure, which is known as the six-helical bundle. Pancera and colleagues' pre-fusion structure shows that HR1 and HR2 are each split up into two smaller helices connected by loops; together, the four helices form a ring encircling the amino and carboxy termini of gp120 (Fig. 1). In turn, these gp120 regions act as a 'safety pin' to prevent gp41 from transiting to the energetically more favoured six-helical-bundle form.

The authors use their structure to make inferences about the conformational changes in Env proteins that take place during fusion, adding detail to the existing model of the process (Fig. 1b). When the cellular receptors are engaged, the safety pin is removed in a two-stage process. First, the top of the trimer opens up. The diminished constraints on the N-terminal segments of gp41 and the space vacated at the trimer axis allow segment B to undergo a loop-to-helix transition. The formation of the resulting long helix (HR1), now



**Figure 1 | A model of HIV-1 fusion to target cells.** HIV-1 Env proteins are trimers of three identical protomers, each with a gp120 and a gp41 subunit. The gp41 subunit comprises cytoplasmic (CT) and transmembrane (TM) domains, and an ectodomain that has six helix-forming segments (A–F), a fusion peptide (FP), a disulphide loop (DSL) and a membrane proximal external region (MPER). **a**, Pancera and colleagues' trimer structure<sup>4</sup> (a single gp140 protomer is shown) contains the gp120 subunit (green) and most of the ectodomain of the gp41 subunit (orange and purple), but omits other gp41 domains. The cysteine amino-acid residues (501C–605C) forming the engineered disulphide bond<sup>12</sup> in the trimer are indicated. **b**, The structure, together with previous

data, helps to build a model of viral fusion to target cells. In the pre-fusion protomer, helix segments A and C, and D and F, are interspersed by loop segments B and E, respectively. On binding to cell-surface receptors, a long helix comprising segments A, B and C forms, punching FP into the host-cell membrane. (The approximate location of the I559P amino-acid substitution, which blocks the loop-to-helix transition in segment B of engineered trimers<sup>11</sup> and thereby stabilizes the pre-fusion structure, is indicated.) A second long helix of segments D, E and F then forms and aligns with the other helix in a hairpin structure. The formation of the trimer of hairpins (called the six-helical bundle) pulls the viral and target-cell membranes together.



containing segments A, B and C, punches the hydrophobic fusion peptide at the N terminus of gp41 into the target-cell membrane. The concomitant removal of one component of the four-helix ring weakens the links between gp120 and gp41. Co-receptor binding then fully removes the safety pin. The gp120 subunit probably now completely dissociates, eliminating any remaining steric constraints on the melding of the viral and cell membranes. In gp41, segments D–F extend to form the second long helix (HR2), and the formation of the six-helical bundle provides sufficient energy to fuse the membranes. Once enough individual trimers (probably around five<sup>5</sup>) have undergone these transitions, the resulting fusion pore in the cell membrane allows the viral core to enter the cell.

The human immune system can prevent HIV-1 from infecting cells by generating neutralizing antibodies that bind to various regions (epitopes) on the pre-fusion Env trimer. But the virus has evolved defences to evade both the generation and the binding of neutralizing antibodies. The most relevant are a dense array of protective glycans, which are less immunogenic than protein surfaces, and the ability to vary the trimer's amino-acid composition. Pancera and colleagues' structure displays just how efficient these obstacles are: one stunning image (see Fig. 6 in the Extended Data) shows the comprehensiveness of the shielding effect of its glycans compared with analogous proteins from influenza virus and respiratory syncytial virus.

Even so, humans can produce antibodies that have extremely broad reactivity against circulating HIV-1 variants. Most known epitopes targeted by such broadly neutralizing antibodies (many of these epitopes include the otherwise protective glycans) are present on the engineered BG505 SOSIP.664 trimers used to generate the existing and new Env structures<sup>2–4,6</sup>. Because inducing such antibodies is a major goal of HIV-1 vaccine development, does this mean that the problem is solved? Unfortunately, no — or, at least, not yet.

There are two main elements in the development of vaccines based on Env: making immunogenic proteins with epitopes that could elicit broadly neutralizing antibodies, and forcing the immune system to respond to them. Current trimer-based immunogens, termed gp140s, are rendered soluble for vaccine development by deleting the transmembrane region of gp41 (ref. 7). However, doing so creates a loose end at the C terminus, which may destabilize inter-subunit interactions<sup>2–4</sup>. For most gp140s, the cleavage site between gp120 and the gp41 ectodomain is deliberately eliminated to 'stabilize' the trimer<sup>7</sup>. However, that stability is illusory, because these 'uncleaved' gp140s predominantly adopt configurations in which semi-dissociated gp120 subunits dangle loosely from the gp41 six-helical bundle<sup>8–10</sup>. The outcome is reminiscent

of the conformational changes outlined above: the dissociation of gp120 from the gp41 ectodomain and the latter's transition to the six-helical bundle post-fusion conformation.

These stability problems, and the resulting loss of epitopes that elicit broadly neutralizing antibodies, are overcome in the BG505 SOSIP.664 trimers through the introduction of an amino-acid substitution (I559P)<sup>11</sup> that prevents the loop-to-helix transition in segment B and of an engineered disulphide bond between cysteine residues 501 and 605 (ref. 12) that pins the gp120 subunits to the four-helix ring of gp41 (Fig. 1). Cleavage of the trimers is also promoted, which seems to strengthen the association between gp120 and gp41 (ref. 9). Together, these engineered changes maintain the soluble trimers in the pre-fusion conformation and preserve key epitopes for eliciting neutralizing antibodies<sup>2–4,6</sup>.

The BG505 SOSIP.664 trimers, which are now vaccine candidates for human trials, are encouragingly immunogenic in animals, provoking a strong and consistent neutralizing-antibody response to the BG505 virus, a neutralization-resistant strain of HIV-1. But they do not elicit broadly neutralizing antibodies. Future strategies to increase the breadth of the antibody response may involve reverse-engineering the immunogens on the basis of antibody evolution<sup>13</sup> and devising different ways of presenting them to the immune system, for example as particulate antigens<sup>14</sup>. Nowadays, structure-guided immunogen

design is the best route to an effective vaccine<sup>15</sup>, and the new structural data will undoubtedly facilitate such improvements. By again placing their stamp on the envelope, the Kwong group has posted a frank warning to the virus. ■

**Rogier W. Sanders and John P. Moore** are in the Department of Microbiology and Immunology, Weill Medical College, Cornell University, New York, New York 10065, USA. R.W.S. is also in the Department of Medical Microbiology, Academic Medical Centre, University of Amsterdam, 1105 AZ Amsterdam, the Netherlands. e-mails: r.w.sanders@amc.uva.nl; jpm2003@med.cornell.edu

1. Kwong, P. D. *et al.* *Nature* **393**, 648–659 (1998).
2. Julien, J.-P. *et al.* *Science* **342**, 1477–1483 (2013).
3. Lyumkis, D. *et al.* *Science* **342**, 1484–1490 (2013).
4. Pancera, M. *et al.* *Nature* **514**, 455–461 (2014).
5. Klasse, P. J. *Virology* **369**, 245–262 (2007).
6. Sanders, R. W. *et al.* *PLoS Pathog.* **9**, e1003618 (2013).
7. Forsell, M. N., Schief, W. R. & Wyatt, R. T. *Curr. Opin. HIV AIDS* **4**, 380–387 (2009).
8. Guttman, M. & Lee, K. K. *J. Virol.* **87**, 11462–11475 (2013).
9. Ringe, R. P. *et al.* *Proc. Natl Acad. Sci. USA* **110**, 18256–18261 (2013).
10. Tran, K. *et al.* *Proc. Natl Acad. Sci. USA* **111**, E738–E747 (2014).
11. Sanders, R. W. *et al.* *J. Virol.* **76**, 8875–8889 (2002).
12. Binley, J. M. *et al.* *J. Virol.* **74**, 627–643 (2000).
13. Medina-Ramírez, M., Sanders, R. W. & Klasse, P. J. *Expert Rev. Vaccines* **13**, 449–452 (2014).
14. Schiller, J. & Chackerian, B. *PLoS Pathog.* **10**, e1004254 (2014).
15. McLellan, J. S. *et al.* *Science* **342**, 592–598 (2013).

This article was published online on 8 October 2014.

#### LUNG DISEASE

## Treatment by cell transplant

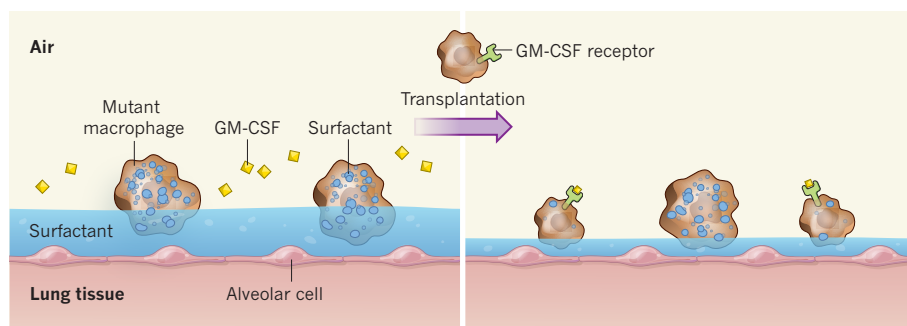
**Transplanting gene-corrected macrophage cells directly into the lungs of mice has been shown to effectively treat their pulmonary alveolar proteinosis, a hereditary lung disease also found in humans. SEE ARTICLE P.450**

MARY JANE THOMASSEN & MANI S. KAVURU

**P**ulmonary alveolar proteinosis (PAP) is a rare lung disease characterized by the accumulation in the lung of white blood cells called alveolar macrophages that are full of surfactant — a compound of phospholipids and proteins that regulates surface tension in the lung — and of vast amounts of extracellular surfactant<sup>1</sup>. Unravelling the cause of this disease, which was first recognized in 1958, is a story that began in 1994 with the serendipitous discovery<sup>2,3</sup> that mice lacking the protein GM-CSF, which is important for macrophage maturation and function, had a mysterious

lung disease that resembled human PAP. In this issue, Suzuki *et al.*<sup>4</sup> (page 450) add a chapter to this story, reporting that transplanting macrophages that correctly respond to GM-CSF into the lungs of mice lacking the GM-CSF receptor effectively treats their disease.

Studies of GM-CSF-deficient mice identified the disease-causing defect as part of the process through which surfactant is broken down by macrophages in the alveolar region of the lung<sup>5</sup> (Fig. 1). And human studies revealed that, although alveolar macrophages from some people with PAP respond to GM-CSF stimulation *in vitro*<sup>6</sup>, the patients express antibodies that neutralize the protein<sup>7</sup>. PAP is



**Figure 1 | Macrophage transplant corrects hereditary PAP in mice.** Macrophages are white blood cells that, among other functions, engulf and destroy cell debris. In the alveolar regions of the lungs, macrophages break down excess surfactant — a compound of phospholipids and proteins that is produced by alveolar cells to reduce surface tension, preventing lung collapse. But mutations that lead to macrophages lacking the receptor for the protein GM-CSF, which is essential for macrophage maturation and function, can cause the build-up of lung surfactant that is characteristic of the disease pulmonary alveolar proteinosis (PAP). Suzuki *et al.*<sup>4</sup> show that transplantation of macrophages expressing the GM-CSF receptor — either macrophages from wild-type mice or mutant macrophages that have been ‘gene-corrected’ *ex vivo* — directly into the lungs of mice treats the disease by allowing adequate surfactant breakdown.

now classified into three forms: autoimmune (acquired), congenital (hereditary) and secondary (linked mainly to cancers of the blood or systemic inflammatory diseases). All forms of PAP are associated with loss of GM-CSF signalling owing either to deficiencies in active GM-CSF or to GM-CSF-receptor mutations, with the exception of some congenital forms that are associated with surfactant-protein abnormalities.

Further work in mouse models and human PAP samples elucidated the steps of surfactant breakdown and the roles of key proteins in alveolar-macrophage biology. These included the transcription factors PU.1, which is involved in the maturation of alveolar macrophages<sup>8</sup>, and PPAR $\gamma$ , which maintains lung homeostasis<sup>9</sup>. Contrary to macrophages in other parts of the body, alveolar macrophages express high levels of PPAR $\gamma$ <sup>9</sup>, suggesting that this protein has a special role in the lung. Indeed, it has been shown that PPAR $\gamma$  is a negative regulator of macrophage activation<sup>10</sup> and that its expression is stimulated by GM-CSF<sup>11</sup>. Alveolar surfactant is 90% lipid, and its catabolism is now known to be regulated by a signalling pathway involving GM-CSF, PPAR $\gamma$  and the protein ABCG1 (refs 12,13).

Therapeutic options for PAP were also first identified in mouse models of the disease. Initial studies showed that GM-CSF-deficient mice could be ‘cured’ either by the administration of exogenous GM-CSF<sup>14</sup> or by overexpression of GM-CSF in epithelial cells of the respiratory tract<sup>15</sup>. These findings led to a trial of treating people who have autoimmune PAP with high doses of GM-CSF, administered under the skin or by inhalation, although a subset of patients failed to respond to this treatment, possibly because of high levels of anti-GM-CSF antibodies in their lungs. An alternative approach to treating the autoimmune form of PAP is to use the monoclonal antibody rituximab, which blocks production

of anti-GM-CSF antibodies<sup>16</sup> and induces increased expression of PPAR $\gamma$  and ABCG1 (ref. 17) in some patients.

However, the only treatment for patients with hereditary PAP is whole-lung lavage (irrigation), which requires general anaesthesia. It has been proposed that hereditary PAP might be corrected by transplantation of healthy bone-marrow cells, which contain stem cells that can differentiate into normal, GM-CSF-sensitive macrophages, and this procedure has indeed been successful in mice<sup>18</sup>. But this approach requires prior myeloablation — the severe or complete depletion of existing bone-marrow cells to avoid rejection of the transplanted cells. Myeloablation is associated with a high risk of infection and death and so bone-marrow transplantation is not routinely performed in patients with PAP.

Suzuki *et al.* sought to design a transplantation approach that circumvents the need for myeloablation. They transplanted macrophages taken from normal mice directly into the lungs of mice deficient for the  $\beta$ -subunit of the GM-CSF receptor (which develop a disorder identical to that in children with hereditary PAP owing to mutations in the receptor’s  $\alpha$ - or  $\beta$ -subunits), and found that this treatment relieved the disease symptoms, normalized the expression of disease-related proteins and extended the lifespan of these mice. The authors then repeated the experiment using macrophages taken from GM-CSF-receptor-deficient mice that had been corrected *ex vivo* (by a process of lentiviral transduction) such that they expressed the  $\beta$ -subunit again, and saw the same effect.

The feasibility of translating pulmonary-macrophage transplantation into a human therapy is strongly supported by these and other recent studies in mice. A previous study<sup>18</sup> had also demonstrated that transplantation of wild-type murine macrophage-progenitor cells into GM-CSF-receptor deficient animals

effectively reduced hereditary PAP disease. Suzuki and colleagues have improved on this approach by using gene-corrected cells from genetically identical animals, thereby avoiding the need for myeloablation and immunosuppression. This bears greater resemblance to the method that would most probably be adopted in humans, which would be to take a patient’s own, mutated, macrophages, correct them *ex vivo*, and return them to the patient. Furthermore, the authors show disease remission with lessening of symptoms in mice.

Suzuki *et al.* also found that, although wild-type bone-marrow-derived macrophages cultured *in vitro* had different characteristics to alveolar macrophages, they adopted a lung-macrophage profile following transplantation into the lung. This result supported earlier work<sup>19</sup> indicating that local microenvironments provide signals that direct macrophage development. Future studies will be needed to determine the optimal dose of transplanted cells for people, the effect of intrinsically elevated GM-CSF levels associated with hereditary PAP and whether additional GM-CSF will need to be administered to promote survival of the transplanted macrophages.

The therapeutic implications of this approach reach beyond the rare disease of hereditary PAP. One can imagine transplantation of autologous gene-corrected macrophages for the treatment of other diseases, such as HIV. Macrophages serve as a reservoir for the virus and individuals lacking a certain macrophage co-receptor are HIV resistant, so transplanting macrophages without the receptor may convey immunity to the virus. Because it seems that local environments provide cues for the development of macrophages with certain characteristics, the possibilities are almost endless once the gene mutations that are relevant to certain disease states are identified. The use of whole-genome sequencing to identify aberrant genes in infants born with life-threatening conditions may further extend the options for macrophage-transplantation therapies<sup>20</sup>. ■

Mary Jane Thomassen is in the Division of Pulmonary, Critical Care and Sleep Medicine, East Carolina University, Greenville, North Carolina 27834, USA. Mani S. Kavuru is in the Division of Pulmonary and Critical Care Medicine, Thomas Jefferson University and Hospital, Philadelphia, Pennsylvania 19107, USA.

e-mails: thomassenm@ecu.edu;  
mani.kavuru@jefferson.edu

- Rosen, S. H., Castleman, B. & Liebow, A. A. *N. Engl. J. Med.* **258**, 1123–1142 (1958).
- Dranoff, G. *et al. Science* **264**, 713–716 (1994).
- Stanley, E. *et al. Proc. Natl Acad. Sci. USA* **91**, 5592–5596 (1994).
- Suzuki, T. *et al. Nature* **514**, 450–454 (2014).
- Yoshida, M. *et al. Am. J. Physiol. Lung Cell Mol. Physiol.* **280**, L379–L386 (2001).
- Thomassen, M. J. *et al. Clin. Immunol.* **95**, 85–92 (2000).



7. Kitamura, T. *et al.* *J. Exp. Med.* **190**, 875–880 (1999).
8. Shibata, Y. *et al.* *Immunity* **15**, 557–567 (2001).
9. Bonfield, T. L. *et al.* *Am. J. Respir. Cell Mol. Biol.* **29**, 677–682 (2003).
10. Ricote, M. *et al.* *Proc. Natl Acad. Sci. USA* **95**, 7614–7619 (1998).
11. Ricote, M. *et al.* *Nature* **391**, 79–82 (1998).
12. Malur, A. *et al.* *Am. J. Physiol. Lung Cell Mol. Physiol.*

- 300**, L73–L80 (2011).
13. Thomassen, M. J. *et al.* *J. Lipid Res.* **48**, 2762–2768 (2007).
14. Huffman, J. A. *et al.* *J. Clin. Invest.* **97**, 649–655 (1996).
15. Reed, J. A. *et al.* *Am. J. Physiol.* **276**, L556–L563 (1999).
16. Kavuru, M. S. *et al.* *Eur. Respir. J.* **38**, 1361–1367 (2011).

17. Malur, A. *et al.* *Respir. Res.* **13**, 46 (2012).
18. Happle, C. *et al.* *Sci. Transl. Med.* **6**, 250ra113 (2014).
19. Guth, A. M. *et al.* *Am. J. Physiol. Lung Cell Mol. Physiol.* **296**, L936–L946 (2009).
20. Saunders, C. J. *et al.* *Sci. Trans. Med.* **4**, 154ra135 (2012).

This article was published online on 1 October 2014.

## ASTRONOMY

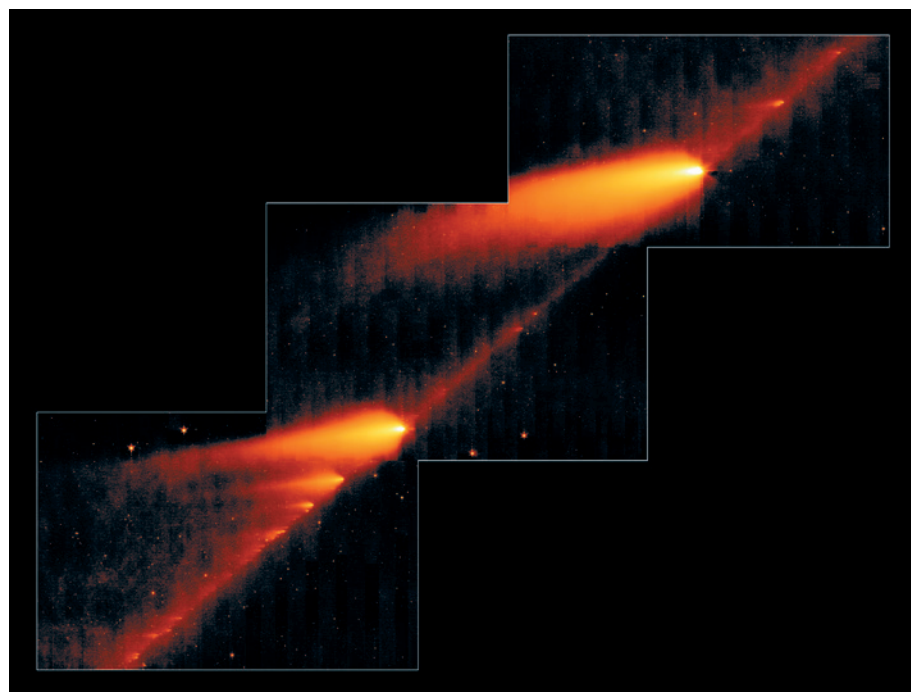
# Hurling comets around a planetary nursery

An analysis of hundreds of star-grazing comets in a young planetary system shows that they form two families: a group of old, dried-out comets and a younger group probably related to the break-up of a larger planetary body. [SEE LETTER P.462](#)

AKI ROBERGE

The Solar System today seems a quiet, orderly place. The planets are on nearly circular orbits, and large collisions with asteroids and comets are rare. But this placid middle age belies the Solar System's turbulent toddler stage nearly 4.6 billion years ago, when asteroids and comets were much more numerous and massive impacts contributed to

the build-up of the terrestrial planets — Mercury, Venus, Earth and Mars. On page 462 of this issue, Kiefer *et al.*<sup>1</sup> open a window on this chaotic late stage of planetary-system formation with 8 years of data on star-grazing comets in the 23-million-year-old planetary system around the star  $\beta$  Pictoris ( $\beta$  Pic). The authors show that the comets in this system are not all alike, but instead consist of two families with distinct dynamical and evaporative properties.



**Figure 1 | A family of fragments from a broken-up comet.** Using 8 years' worth of data on the 23-million-year-old planetary system around the star  $\beta$  Pictoris, Kiefer *et al.*<sup>1</sup> discovered nearly 500 individual comets passing between the star and our vantage point on Earth. They found a family of comets in this sample that approached the star from a single direction and that probably came from the break-up of a larger icy planetary body. This image of the Solar System comet 73P/Schwassmann–Wachmann 3 indicates that it has been broken up into a 'string of pearls' — a similar phenomenon to that discovered by Kiefer and colleagues but on a much smaller scale.

The  $\beta$  Pic planetary system is special. Among its notable features are its young age<sup>2</sup>; its disk of gas and dust originating from the destruction of comets and asteroids<sup>3</sup>; its giant planet,  $\beta$  Pictoris b, one of the first directly imaged exoplanets<sup>4</sup>; and its huge clumps of orbiting carbon monoxide, which indicate the presence of another giant planet far from the star<sup>5</sup>. Objects called falling evaporating bodies (FEBs) were also found early on<sup>6</sup>. Like a Solar System comet (Fig. 1), an FEB heats up as it approaches the star and starts to evaporate, releasing gas and dust. When one of these bodies passes through Earth's line of sight to the star (a transit), the gas absorbs some portion of the starlight, producing time-variable absorption features in the stellar spectra. Each of these features indicates the transit of a star-grazing comet. This phenomenon occurs in the present-day Solar System, but it happens much more frequently around  $\beta$  Pic.

Kiefer *et al.* proved this with their large collection of optical spectra of  $\beta$  Pic obtained using the HARPS instrument on the European Southern Observatory's 3.6-metre telescope in La Silla, Chile. This instrument is more typically used to discover exoplanets using the radial-velocity technique, which measures the change in the velocity of a star caused by the gravitational pull of orbiting planets. But it is also well suited to detecting star-grazing comets through changes in the absorption lines of ionized calcium gas in the stellar spectra. In their 8 years of HARPS data on  $\beta$  Pic, Kiefer and colleagues confirmed 493 individual comets.

This large set of comets allowed the discovery of distinct families within the population. Measurements of the comet radial velocities relative to the star revealed two groups, one with a broad distribution of velocities (population S) and the other with a distribution that narrowly peaked at around 15 kilometres per second (population D). This means that the D comets are all approaching the star from a particular direction, whereas the S comets approach from a wide range of directions. The two populations had other distinct dynamical properties indicating that the D comets transit the star at greater distances than the S comets.

Previous studies of the  $\beta$  Pic FEBs had already hinted at this dynamical information<sup>7</sup>, but Kiefer and colleagues also measured the evaporative properties of the star-grazing comets. Because comets evaporate faster the hotter and closer they are to a star, one would expect the S comets to produce more gas than the D comets, which are farther from the star

when they transit. But the opposite is true. The authors used their data to calculate the evaporation efficiency for each comet, and showed that the D comets generate more gas than the S comets. The former are either physically larger than the latter or they are 'fresher', with more exposed ices at their surfaces that can evaporate. The D comets' characteristics are consistent with the idea that the comets are the fragments of a larger icy planetary body that recently broke up. Further work will be needed to determine exactly how recently.

There is ambiguity about whether the

D comets are fresher or simply larger than the S comets. But either way, they are distinctly different in both their dynamics and their physical properties and hold clues to the extreme collisions occurring in the  $\beta$  Pic planetary system. Kiefer and colleagues' work adds to the transformation in our understanding of planet formation that has developed over the past three decades, from a slow, stately process to a dynamic and sometimes violent one. ■

**Aki Roberge is at the NASA Goddard Space Flight Center, Exoplanets and Stellar**

**Astrophysics Laboratory, Greenbelt, Maryland 20771, USA.**

**e-mail: aki.roberge@nasa.gov**

1. Kiefer, F. *et al. Nature* **514**, 462–464 (2014).
2. Mamajek, E. E. & Bell, C. P. M. *Mon. Not. R. Astron. Soc.* (in the press); preprint at <http://arxiv.org/abs/1409.2737>.
3. Smith, B. A. & Terrile, R. J. *Science* **226**, 1421–1424 (1984).
4. Lagrange, A.-M. *et al. Science* **329**, 57–59 (2010).
5. Dent, W. R. F. *et al. Science* **343**, 1490–1492 (2014).
6. Lagrange, A. M., Ferlet, R. & Vidal-Madjar, A. *Astron. Astrophys.* **173**, 289–292 (1987).
7. Beust, H., Vidal-Madjar, A., Ferlet, R. & Lagrange-Henri, A. M. *Astron. Astrophys.* **236**, 202–216 (1990).

## PHYSIOLOGY

# Relax and come in

**During inflammation, lymph nodes swell with an influx of immune cells. New findings identify a signalling pathway that induces relaxation in the contractile cells that give structure to these organs. SEE LETTER P.498**

KARI VAAHTOMERI & MICHAEL SIXT

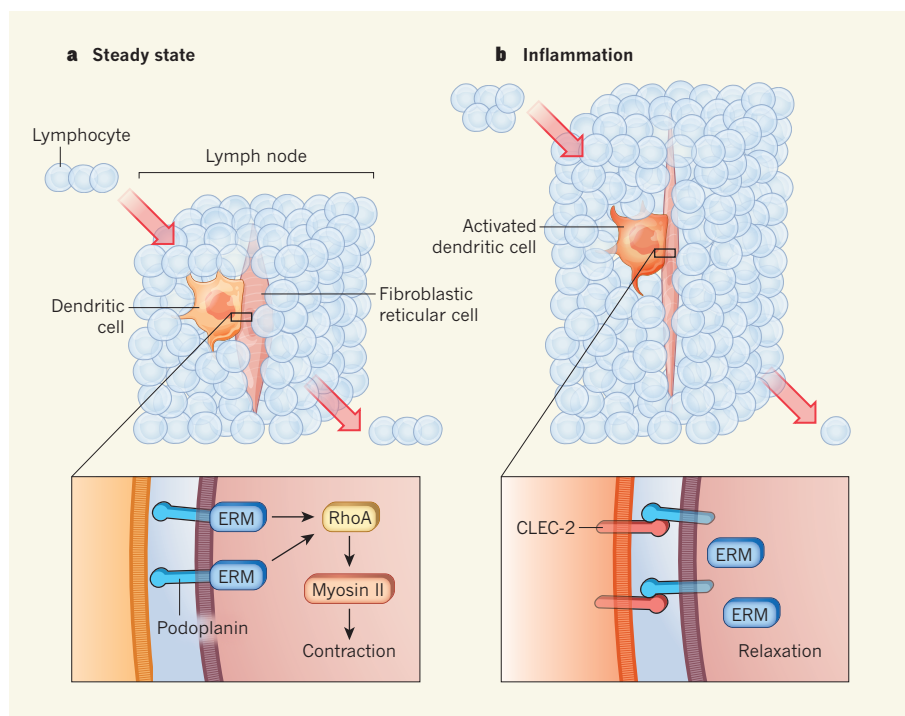
Mechanical forces are key elements in the developmental control of organ size<sup>1</sup>. Adjusting the size of lymph nodes is a special challenge, because immune cells continually enter and exit the organs from the bloodstream, and the number of cells the nodes contain can rapidly increase during an immune or inflammatory response. Despite these dramatic changes in cellularity, the structural backbone of lymph nodes — a network of stromal fibroblastic reticular cells — remains relatively stable<sup>2</sup>. In this issue, Acton *et al.*<sup>3</sup> (page 498) show that, during inflammation, immune cells called dendritic cells transmit a signal that triggers the physical relaxation of these stromal cells, thereby making space for more immune cells to enter the lymph nodes. This mechanical control aids rapid swelling of the organ, without substantial remodelling or proliferation of the stromal cells.

Lymph nodes are the sites of key immune-cell interactions. In these organs, dendritic cells (DCs) present pathogen molecules (antigens) carried in from peripheral tissues to lymphocytes (particularly T cells), which travel between lymph nodes through the blood and lymphatic circulation. In the lymph nodes, T cells rapidly fan out to scan the DCs for antigens that match their specific T-cell receptors. Although it is known that the same guidance cues attract DCs and T cells to a shared lymph-node compartment, it is unclear how the entry and exit of T cells is coordinated quantitatively. It has been suggested<sup>4</sup> that, on exiting the bloodstream, T cells remain in a waiting position until there is enough space to enter the lymph node. Such mechanical regulation might balance influx and efflux, but this must

shift during inflammatory states to allow for swelling of the organ and intensified scanning of DCs by T cells.

The network of fibroblastic reticular cells

(FRCs) permeates the lymph node like the skeleton of a sponge. FRCs are the main producers of chemokines, the factors that attract DCs and T cells and keep them motile. The migrating cells also use the adhesive FRC surface as a guidance structure. In addition to these roles in orchestrating immune-cell traffic, the FRC network forms an interconnected system of micro-scale conduits, which are connected to the lymphatic system and bloodstream and serve as a system for distributing small molecules to resident DCs. Although FRCs have many characteristics of epithelial cells (which line the body's surfaces and cavities), they also express smooth-muscle actin<sup>5</sup>,



**Figure 1 | Lymph-node swelling during the inflammatory response.** Immune cells, including lymphocytes and dendritic cells (DCs), continually enter and exit lymph nodes. The number of cells that the node can accommodate is restricted by the available space, which Acton *et al.*<sup>3</sup> show is regulated by the contractile state of the fibroblastic reticular cells (FRCs) that form a network providing lymph-node structure. **a**, In the steady state, the FRCs are kept in a contracted state by activity of the transmembrane protein podoplanin, which signals through ERM proteins to RhoA and GEF-H1 (not shown), inducing myosin-II-mediated contraction of the actomyosin protein network. **b**, During inflammation, activation of DCs induces expression of the surface molecule CLEC-2, which binds podoplanin and switches off this signalling pathway. This leads to relaxation of the FRC network, allowing the lymph node to accommodate more lymphocytes and to swell.



a protein normally found in cell types that can physically contract.

Acton *et al.* show that FRCs do have contractile function and that they use this to tune the tension of the lymph node, demonstrating that it is not the capsule tissue surrounding the organ that restricts its size, but its internal scaffold. The authors investigated signalling in these cells through podoplanin, an FRC transmembrane protein that binds the receptor CLEC-2, which is induced on DCs following contact with pathogens. They found that, in the absence of CLEC-2 as a ligand, podoplanin transmits signals through proteins of the ERM family to the proteins GEF-H1 and RhoA, which regulate actin contractility mediated by the protein myosin II. This signalling maintains FRCs in a highly contractile state. Binding of podoplanin by CLEC-2 caused podoplanin to redistribute to another membrane compartment, where it stopped signalling to ERM proteins and thereby relaxed actomyosin contraction. This switched the cells from a contractile state dominated by actomyosin filaments to a protrusive, relaxed and elongated form.

These *in vitro* findings suggested a plausible scenario for events *in vivo* (Fig. 1). Activated DCs in the lymph nodes — either DCs that are activated in the periphery and migrate into the lymph node, or DCs resident in the lymph node that are activated by factors carried by the FRC conduit system — are induced to upregulate CLEC-2, causing relaxation of the FRCs. This loosens up the lymph-node structure, allowing influx of additional T cells and effectively expanding the volume of the organ.

In line with this model and previous studies<sup>2</sup>, Acton and colleagues found that, at early time points after an inflammatory stimulus in mice, when the cellularity and volume of their lymph nodes has already tripled, FRCs did not undergo significant proliferation. Instead, the spacing of the FRC network increased, indicating a stretched configuration of the cells. To test the potential involvement of podoplanin signalling in this process, the authors studied mice that were lacking CLEC-2 in DCs, and found that they show severely impaired lymph-node swelling after immunization. However, swelling was restored when podoplanin on FRCs was artificially bound by injecting the mice with CLEC-2 protein.

This newly identified mechanism for lymph-node relaxation is thought-provoking. Besides lymph nodes, podoplanin-expressing FRC-like cellular networks are found in the thymus, spleen and tumour-associated stroma, and may play a similar part in regulating the size and cellularity of these organs and tissues. If the size of the local immune compartment could be pharmacologically altered by tuning the contractile state, such networks might even serve as potential targets for immunomodulation. Podoplanin is also induced on most tissue fibroblast cells during inflammatory states<sup>6</sup>. In this situation, CLEC-2–podoplanin signalling

might biomechanically resolve swelling and inflammation by squeezing the tissue once CLEC-2 levels drop with the disappearance of pathogenic stimuli.

Although it is highly plausible that DCs act as key messengers to mediate lymph-node relaxation, it seems probable that FRC contractility can also be tuned by other input signals. In this context, it is interesting to note that in many mammals (horses, dogs, humans and especially deep-diving seals), the spleen can contract under low-oxygen conditions to expel an 'emergency reservoir' of oxygenated red blood cells<sup>7</sup>. Furthermore, elevated numbers of white blood cells have been associated with spleen and lymph-node contractions, indicating that these responses, which are

mainly triggered by nervous-system signals<sup>8</sup>, could also serve immunological functions. ■

**Kari Vaahtomeri and Michael Sixt** are at the *Institute of Science and Technology (IST) Austria*, 3400 Klosterneuburg, Austria.  
e-mail: michael.sixt@ist.ac.at

1. Low, B. C. *et al.* *FEBS Lett.* **588**, 2663–2670 (2014).
2. Yang, C.-Y. *et al.* *Proc. Natl Acad. Sci. USA* **111**, E109–E118 (2014).
3. Acton, S. E. *et al.* *Nature* **514**, 498–502 (2014).
4. Mionnet, C. *et al.* *Blood* **118**, 6115–6122 (2011).
5. Luther, S. A., Vogt, T. K. & Siegfert, S. *Immunol. Lett.* **138**, 9–11 (2011).
6. Peduto, L. *et al.* *J. Immunol.* **182**, 5789–5799 (2009).
7. Cabanac, A., Folkow, L. P. & Blix, A. S. *J. Appl. Physiol.* **82**, 1989–1994 (1997).
8. McHale, N. G. & Thornbury, K. D. *Exp. Physiol.* **75**, 847–850 (1990).

## HIV

## Potency needs constancy

**In a finding that highlights ways to optimize the efficacy of antibody-based therapeutics and vaccines, the activity of potent HIV-1-neutralizing antibodies has been confirmed to depend on cellular binding to the antibodies' Fc regions.**

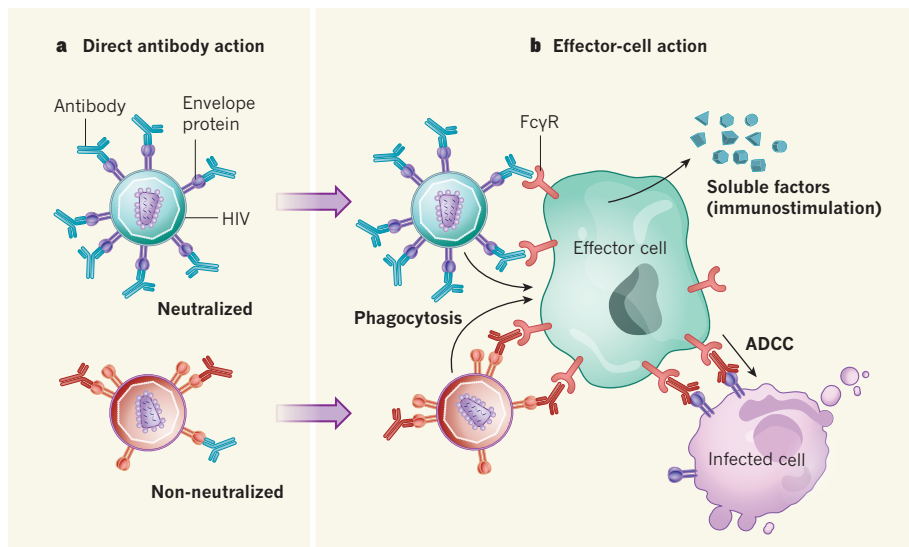
ALEXANDRA TRKOLA

Protection conferred against virus infection by licensed antiviral vaccines is mostly due to the generation of 'neutralizing' antibodies that bind to virus particles in such a way that they block their entry into a cell<sup>1</sup>. Current strategies for designing treatments and vaccines for HIV-1 focus on such antibodies, particularly on the highly potent, broadly neutralizing antibodies that are produced during the immune responses of rare infected individuals<sup>2,3</sup> and which bind to the envelope protein (a highly variable protein) of multiple strains of the virus. However, it is becoming clear that the impact of these antibodies is not limited to direct viral neutralization through binding. Writing in *Cell*, Bournazos *et al.*<sup>4</sup> show that *in vivo* activity of even the most potent broadly neutralizing anti-HIV-1 antibodies relies on antiviral functions that are triggered by binding of cellular receptors — the Fcγ receptors — to the 'tail' region of antibodies of the immunoglobulin G class, binding that was previously found<sup>5</sup> to be crucial for sustaining the *in vivo* activity of less-potent neutralizing antibodies.

Despite showing promising neutralizing activity *in vitro*, attempts to use broadly neutralizing antibodies to control HIV-1 infections *in vivo* have had limited success. Up to now, only the most potent neutralizing antibodies, used at high dose and often only

in combination, have yielded measurable control of HIV-1, and this effect rapidly declines with waning antibody levels and as a result of mutations that allow the virus to 'escape' antibody binding<sup>6–9</sup>. Difficulties with tissue penetration<sup>10</sup> and reduced neutralizing activity during cell-to-cell viral transmission<sup>11</sup> may also contribute to the lower efficacy of these antibodies *in vivo* and necessitate the maintenance of antibody levels orders of magnitude above those sufficient *in vitro*<sup>6–9</sup>.

Antibodies are notable for their capacity to link the adaptive and innate arms of the immune response: the variable domains of an antibody specifically recognize different targets, and their constant tail domain, the Fc region, stimulates an array of immune defences. The Fc region is bound by proteins of the complement system, which activate signalling pathways to induce inflammatory responses and destruction of antibody-bound cells or viruses. The Fc region of immunoglobulin G is also recognized by activating and inhibitory Fcγ receptors (FcγRs) on effector cells of the immune system, such as monocytes, macrophages, dendritic cells, neutrophils and natural killer cells. Fc binding provides these cells with essential stimulatory and regulatory signals<sup>12</sup>. FcγR-mediated effector-cell functions include antibody-dependent cellular cytotoxicity (in which antibodies bind to viral proteins expressed on the infected cell, triggering its killing by effector cells), phagocytic



**Figure 1 | Antibody-mediated viral control.** **a**, Binding of certain (blue) but not other (red) antibodies' variable regions to an HIV-1 virus particle can directly neutralize the virus if all envelope proteins required for cell entry are inhibited. **b**, However, Bournazos *et al.*<sup>4</sup> show that the antiviral activity of these neutralizing antibodies *in vivo* depends on effects elicited when the constant (Fc) regions of immunoglobulin G antibody molecules are bound by Fcγ receptors (FcγRs) on the surface of effector cells of the immune system. These effects include engulfment and destruction (phagocytosis) of antibody-bound virus particles (including neutralized and non-neutralized particles); the secretion of soluble factors that stimulate other immune activities; and direct killing (antibody-dependent cellular cytotoxicity; ADCC) of virus-infected cells.

clearance (cellular engulfment and destruction of viruses), and the release of soluble antimicrobial and immunomodulatory factors, such as cytokines and chemokines (Fig. 1).

Recruitment of effector functions has long been a focus in the harnessing of antibody-based protection against HIV-1<sup>4,5,12–14</sup>, but the precise contributions of direct antibody-mediated virus neutralization, activity of the complement system and FcγR-mediated effector functions remain unresolved<sup>13</sup>. Pioneering work indicated that neutralizing antibodies require FcγR functions for their *in vivo* activity<sup>5</sup>, whereas non-neutralizing antibodies that act solely through effector functions have shown limited *in vivo* activity against HIV-1<sup>13,15</sup>, suggesting that a combination of both neutralization and effector-mediated activity is needed. By contrast, complement action seems dispensable, at least when neutralizing antibodies are used as a pre-exposure prophylaxis (to try to prevent infection)<sup>5</sup>. Now, Bournazos *et al.* expand on the previous finding that FcγR-dependent mechanisms are required<sup>5</sup> by investigating the effect of antibody treatment on HIV-1 infection in mice in the presence or absence of Fc–FcγR interactions, either by modulating the antibodies' Fc region or by using mice lacking FcγRs.

Particularly notable is the authors' finding that, despite their superior potency, even the action of broadly neutralizing antibodies is largely dependent on interactions with activating FcγRs. When the researchers engineered the antibodies to improve the strength of Fc–FcγR binding, they observed increased viral control *in vivo*, highlighting the potential of antibody improvement for therapeutic

use. On reflection, however, this dependence on FcγR functions might prove to be a crucial limitation on the use of engineered molecules designed to inhibit HIV entry to cells by targeting viral envelope proteins, because such molecules lack the ability to recruit immune effector functions.

Intriguingly, in experiments using HIV particles that can complete only one round of infection, such that newly infected cells do not express HIV envelope proteins and hence lack binding sites for antibodies, Bournazos *et al.* still saw FcγR-dependent antibody-mediated viral control. This suggests that classical antibody-dependent cellular cytotoxicity can be ruled out as a sole driving force, and that effector-cell release of soluble antiviral factors and the phagocytic clearance of viral particles may be decisive.

Nevertheless, a central question that remains is, why would a potent neutralizing antibody need to rely on fast, FcγR-mediated removal of antibody-bound viral particles? There are several possible explanations. Neutralization can be reversible if the antibody binding has a high off-rate<sup>16–18</sup>, and if an antibody fails to irreversibly neutralize the virus, it may be crucial to eliminate these viruses rapidly before they regain infectivity. Tissue penetration may be another limiting factor<sup>10</sup>: lower antibody doses at certain sites may mean that not all envelope proteins on the virus are immediately bound and neutralized.

How many envelope proteins HIV carries, how many of these are needed to infect (and in turn must be neutralized), and how many antibody molecules are needed to trigger irreversible neutralization of envelope proteins

will all also contribute to the rate of virus inactivation<sup>19</sup>. Alongside these stoichiometric requirements, the kinetics of neutralization will be steered by the on-rate of antibody binding. Considering all these factors, enhanced clearance of antibody-bound viral particles may become relevant in situations in which the virus is not fully neutralized, either because the threshold of neutralizing-antibody binding required for inactivation has not been reached, or because the specific neutralizing antibody fails to irreversibly block the virus. In support of the idea that quantity of antibody, and hence antibody occupancy, plays a part, Bournazos and colleagues observed that eliciting FcγR effector functions had an additive effect on viral control at lower, but not higher, antibody doses.

Further work will be needed to precisely quantify the impact of phagocytic clearance, and to define if only one or a combination of FcγR-mediated functions is needed to achieve *in vivo* control of HIV-1 by neutralizing antibodies. Several technical challenges may arise in attempts to investigate this. HIV-1 infection is commonly monitored by quantifying levels of viral RNA, but this approach does not assess the infectivity of the viral particles present, and so both neutralized and non-neutralized virions will be counted. Thus, short-term experiments in which the effects of neutralizing antibodies are assessed solely on the basis of a reduction in RNA levels will not be able to quantify the contribution of Fc-mediated effects. Long-term monitoring of infections, as performed by Bournazos *et al.*<sup>4</sup> and in previous work<sup>5</sup>, in combination with measurements of circulating infectious virus and infected cells, will be key to quantifying the influence of FcγR-dependent mechanisms. ■

Alexandra Trkola is at the Institute of Medical Virology, University of Zurich, 8057 Zurich, Switzerland.  
e-mail: trkola.alexandra@virology.uzh.ch

1. Koff, W. C. *et al.* *Science* **340**, 1232910 (2013).
2. Burton, D. R. *et al.* *Cell Host Microbe* **12**, 396–407 (2012).
3. Mascola, J. R. & Montefiori, D. C. *Annu. Rev. Immunol.* **28**, 413–444 (2010).
4. Bournazos, S. *et al.* *Cell* **158**, 1243–1253 (2014).
5. Hessel, A. J. *et al.* *Nature* **449**, 101–104 (2007).
6. Barouch, D. H. *et al.* *Nature* **503**, 224–228 (2013).
7. Klein, F. *et al.* *Nature* **492**, 118–122 (2012).
8. Trkola, A. *et al.* *Nature Med.* **11**, 615–622 (2005).
9. Poignard, P. *et al.* *Immunity* **10**, 431–438 (1999).
10. Ko, S.-Y. *et al.* *Nature* <http://dx.doi.org/10.1038/nature13612> (2014).
11. Abela, I. A. *et al.* *PLoS Pathog.* **8**, e1002634 (2012).
12. Ackerman, M. E., Dugast, A. S. & Alter, G. *Annu. Rev. Med.* **63**, 113–130 (2012).
13. Ackerman, M. E. & Alter, G. *Curr. HIV Res.* **11**, 365–377 (2013).
14. Huber, M. *et al.* *PLoS Med.* **3**, e441 (2006).
15. Burton, D. R. *et al.* *Proc. Natl Acad. Sci. USA* **108**, 11181–11186 (2011).
16. Platt, E. J., Gomes, M. M. & Kabat, D. *Proc. Natl Acad. Sci. USA* **109**, 7829–7834 (2012).
17. Rupprecht, C. R. *et al.* *J. Exp. Med.* **208**, 439–454 (2011).
18. Yasmeen, A. *et al.* *Retrovirology* **11**, 41 (2014).
19. Magnus, C., Rusert, P., Bonhoeffer, S., Trkola, A. & Regoes, R. R. *J. Virol.* **83**, 1523–1531 (2009).



# Genome sequence of a 45,000-year-old modern human from western Siberia

Qiaomei Fu<sup>1,2</sup>, Heng Li<sup>3,4</sup>, Priya Moorjani<sup>3,5</sup>, Flora Jay<sup>6</sup>, Sergey M. Slepchenko<sup>7</sup>, Aleksei A. Bondarev<sup>8</sup>, Philip L. F. Johnson<sup>9</sup>, Ayinuer Aximu-Petri<sup>2</sup>, Kay Prüfer<sup>2</sup>, Cesare de Filippo<sup>2</sup>, Matthias Meyer<sup>2</sup>, Nicolas Zwyns<sup>10,11</sup>, Domingo C. Salazar-García<sup>10,12,13,14</sup>, Yaroslav V. Kuzmin<sup>15</sup>, Susan G. Keates<sup>15</sup>, Pavel A. Kosintsev<sup>16</sup>, Dmitry I. Razhev<sup>7</sup>, Michael P. Richards<sup>10,17</sup>, Nikolai V. Peristov<sup>18</sup>, Michael Lachmann<sup>2,19</sup>, Katerina Douka<sup>20</sup>, Thomas F. G. Higham<sup>20</sup>, Montgomery Slatkin<sup>6</sup>, Jean-Jacques Hublin<sup>10</sup>, David Reich<sup>3,4,21</sup>, Janet Kelso<sup>2</sup>, T. Bence Viola<sup>2,10</sup> & Svante Pääbo<sup>2</sup>

**We present the high-quality genome sequence of a ~45,000-year-old modern human male from Siberia. This individual derives from a population that lived before—or simultaneously with—the separation of the populations in western and eastern Eurasia and carries a similar amount of Neanderthal ancestry as present-day Eurasians. However, the genomic segments of Neanderthal ancestry are substantially longer than those observed in present-day individuals, indicating that Neanderthal gene flow into the ancestors of this individual occurred 7,000–13,000 years before he lived. We estimate an autosomal mutation rate of  $0.4 \times 10^{-9}$  to  $0.6 \times 10^{-9}$  per site per year, a Y chromosomal mutation rate of  $0.7 \times 10^{-9}$  to  $0.9 \times 10^{-9}$  per site per year based on the additional substitutions that have occurred in present-day non-Africans compared to this genome, and a mitochondrial mutation rate of  $1.8 \times 10^{-8}$  to  $3.2 \times 10^{-8}$  per site per year based on the age of the bone.**

In 2008, a relatively complete left human femoral diaphysis was discovered on the banks of the river Irtysh (Fig. 1a, c, d), near the settlement of Ust'-Ishim in western Siberia (Omsk Oblast, Russian Federation). Although the exact locality is unclear, the femur was eroding out of alluvial deposits on the left bank of the river, north of Ust'-Ishim. Here, Late Pleistocene and probably redeposited Middle Pleistocene fossils are found in sand and gravel layers that are about 50,000–30,000 years old (that is, from Marine Oxygen Isotope Stage 3).

## Morphology, dating and diet

The proximal end of the bone shows a large gluteal buttress and gluteal tuberosity, while the midshaft is dominated by a marked *linea aspera*, resulting in a teardrop-shaped cross-section (Fig. 1e, f) (for details, see Supplementary Information section 3). The morphology of the proximal end of the shaft is similar to Upper Paleolithic modern humans and distinct from Neanderthals (Supplementary Table 3.1, Supplementary Fig. 3.2.), while the teardrop-shaped cross section of the midshaft is similar to most Upper Paleolithic humans and early anatomically modern humans<sup>1</sup>. Taken together, this suggests that the Ust'-Ishim femur derives from a modern human.

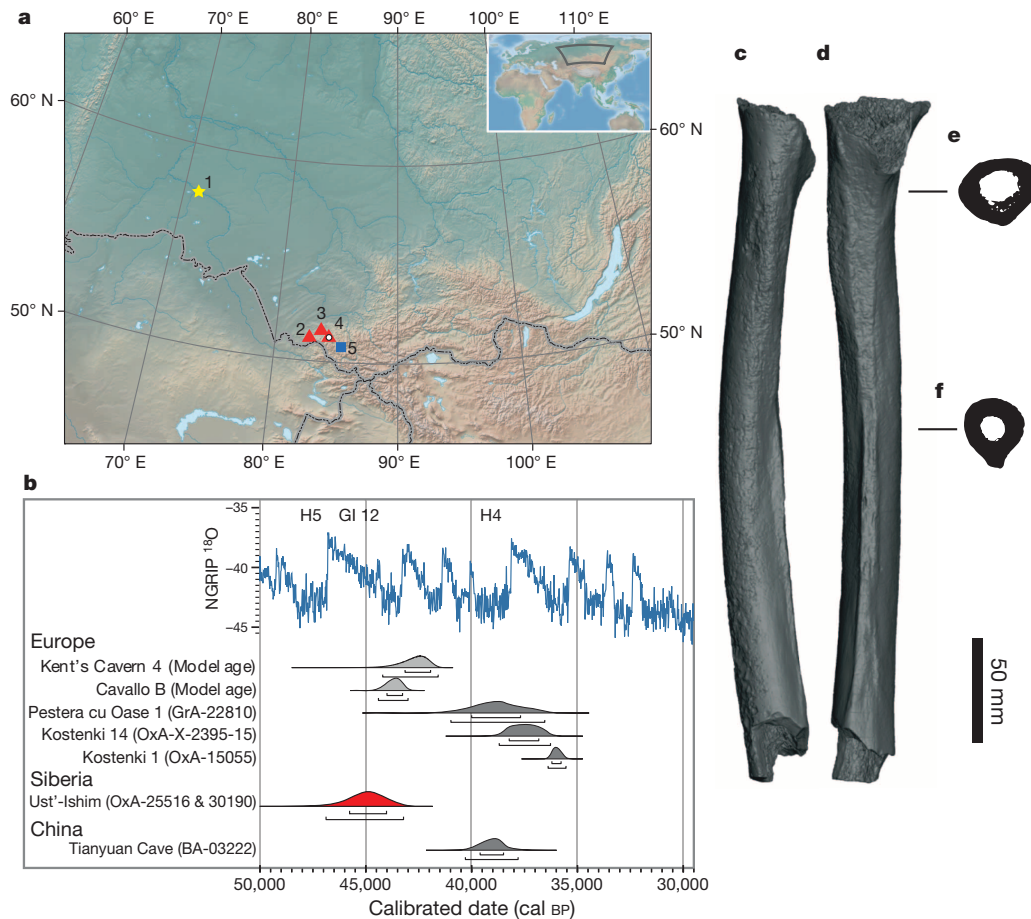
Two samples of 890 mg and 450 mg of the bone were removed on separate occasions for dating. Collagen preservation satisfied all criteria for dating<sup>2</sup> and after ultrafiltration we obtained ages of  $41,400 \pm 1,300$  years before present (BP) (OxA-25516) and  $41,400 \pm 1,400$  BP (OxA-30190). These two dates, when combined and corrected for fluctuations of atmospheric <sup>14</sup>C through time, correspond to an age of about 45,000 calibrated

years BP (46,880–43,210 cal BP at 95.4% probability, Supplementary Information section 1). The Ust'-Ishim individual is therefore the oldest directly radiocarbon-dated modern human outside Africa and the Middle East (Fig. 1b). Carbon and nitrogen isotope ratios indicate that the diet of the Ust'-Ishim individual (Supplementary Information section 4) was based on terrestrial C<sub>3</sub> plants and animals that consumed them, but also that an important part of his dietary protein may have come from aquatic foods, probably freshwater fish, something that has been observed in other early Upper Palaeolithic humans from Europe<sup>3</sup>.

## DNA retrieval and sequencing

Nine samples of between 41 and 130 mg of bone material were removed from the distal part of the femur and used to construct DNA libraries using a protocol designed to facilitate the retrieval of short and damaged DNA<sup>4</sup>. The percentage of DNA fragments in these libraries that could be mapped to the human genome varied between 1.8% and 10.0% (Supplementary Table 1.1). From the extract containing the highest proportion of human DNA, eight further libraries were constructed. Each of these libraries was treated with uracil-DNA glycosylase and endonuclease VIII to remove deaminated cytosine residues, and library molecules with inserts shorter than approximately 35 base pairs (bp) were depleted by preparative acrylamide gel electrophoresis before sequencing on the Illumina HiSeq platform (Supplementary Information section 6). In total, 42-fold sequence coverage of the ~1.86 gigabases (Gb) of the autosomal genome to which short fragments can be confidently mapped was generated. The coverage of the X and Y chromosomes was approximately

<sup>1</sup>Key Laboratory of Vertebrate Evolution and Human Origins of Chinese Academy of Sciences, IVPP, CAS, Beijing 100044, China. <sup>2</sup>Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, D-04103 Leipzig, Germany. <sup>3</sup>Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. <sup>4</sup>Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>5</sup>Department of Biological Sciences, Columbia University, New York, New York 10027, USA. <sup>6</sup>Department of Integrative Biology, University of California, Berkeley, California 94720-3140, USA. <sup>7</sup>Institute for Problems of the Development of the North, Siberian Branch of the Russian Academy of Sciences, Tyumen 625026, Russia. <sup>8</sup>Expert Criminalistics Center, Omsk Division of the Ministry of Internal Affairs, Omsk 644007, Russia. <sup>9</sup>Department of Biology, Emory University, Atlanta, Georgia 30322, USA. <sup>10</sup>Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology, D-04103 Leipzig, Germany. <sup>11</sup>Department of Anthropology, University of California, Davis, California 95616, USA. <sup>12</sup>Department of Archaeology, University of Cape Town, Cape Town 7701, South Africa. <sup>13</sup>Departament de Prehistòria i Arqueologia, Universitat de València, Valencia 46101, Spain. <sup>14</sup>Research Group on Plant Foods in Hominin Dietary Ecology, Max-Planck Institute for Evolutionary Anthropology, D-04103 Leipzig, Germany. <sup>15</sup>Institute of Geology and Mineralogy, Siberian Branch of the Russian Academy of Sciences, Novosibirsk 630090, Russia. <sup>16</sup>Institute of Plant and Animal Ecology, Urals Branch of the Russian Academy of Sciences, Yekaterinburg 620144, Russia. <sup>17</sup>Laboratory of Archaeology, Department of Anthropology, University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada. <sup>18</sup>Siberian Cultural Center, Omsk 644010, Russia. <sup>19</sup>Santa Fe Institute, Santa Fe, New Mexico 87501, USA. <sup>20</sup>Oxford Radiocarbon Accelerator Unit, Research Laboratory for Archaeology and the History of Art, University of Oxford, Oxford OX1 3QY, UK. <sup>21</sup>Howard Hughes Medical Institute, Harvard Medical School, Boston, Massachusetts 02115, USA.



**Figure 1 | Geographic location, morphology and dating.** **a**, Map of Siberia with major archaeological sites. Red triangles: Neanderthal fossils; white circle within a red (Neanderthal) triangle: Denisovan fossils; blue square: Initial Upper Palaeolithic sites; yellow asterisk: Ust'-Ishim. 1: Ust'-Ishim; 2: Chagyrskaya Cave; 3: Okladnikov Cave; 4: Denisova Cave; 5: Kara-Bom. **b**, Radiocarbon ages of early modern human fossils in northern Eurasia and the NGRIP  $\delta^{18}\text{O}$  palaeotemperature record. Specimens in light grey are indirectly

dated (OxCal v4.2.3 (ref. 33); r:5 IntCal13 atmospheric curve<sup>34</sup>). H5: Heinrich 5 event, H4: Heinrich 4 event, GI 12: Greenland Interstadial 12. For a more extensive comparison see Supplementary Information Fig. 2.1. **c–f**, The Ust'-Ishim 1 femur. **c**, Lateral view. **d**, Posterior view. **e**, Cross-section at the 80 percent level. **f**, Cross-section at the midshaft. For other views see Supplementary Fig. 3.1.

half that of the autosomes ( $\sim 22$ -fold), indicating that the bone comes from a male. A likelihood method estimated present-day human mitochondrial DNA (mtDNA) contamination<sup>5</sup> to 0.50% (95% confidence interval (CI) 0.26–0.94%), whereas a method that uses the frequency of non-consensus bases in autosomal sequences estimated the contamination to be less than 0.13% (Supplementary Information section 7). Thus, less than 1% of the hominin DNA fragments sequenced are estimated to be extraneous to the bone. After consensus genotype calling, such low levels of contamination will tend to be eliminated.

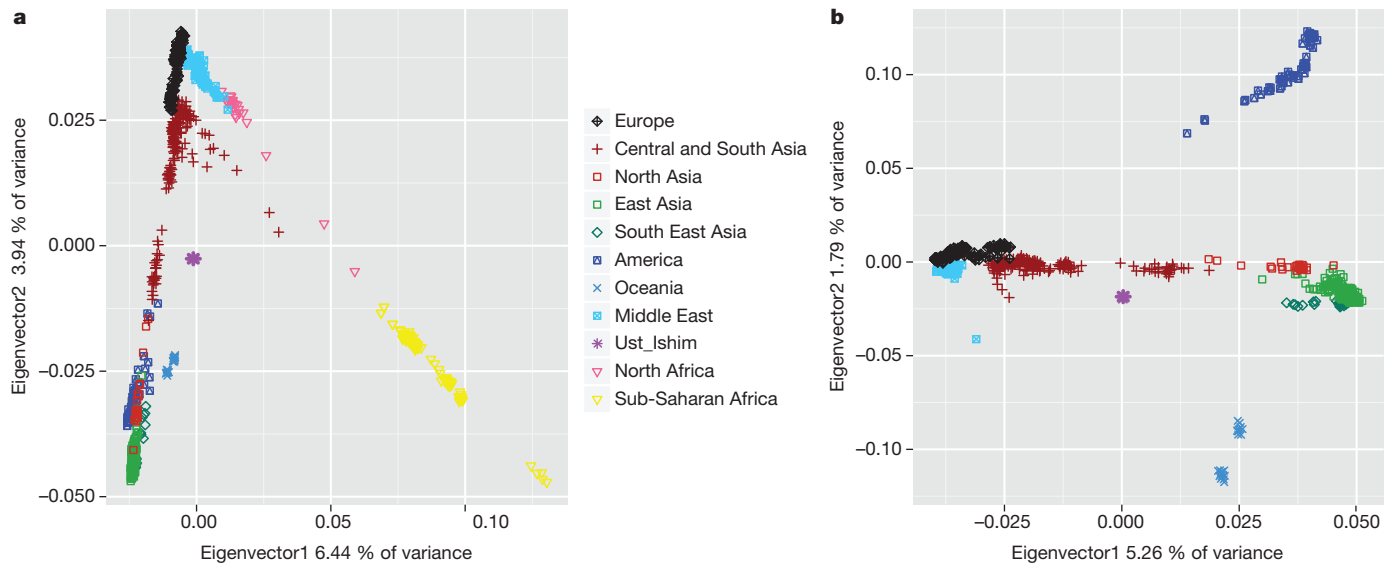
### Population relationships

About 7.7 positions per 10,000 are heterozygous in the Ust'-Ishim genome, whereas between 9.6 and 10.5 positions are heterozygous in present-day Africans and 5.5 and 7.7 in present-day non-Africans (Supplementary Information section 12). Thus, with respect to genetic diversity, the population to which the Ust'-Ishim individual belonged was more similar to present-day Eurasians than to present-day Africans, which probably reflects the out-of-Africa bottleneck shared by non-African populations. The Ust'-Ishim mtDNA sequence falls at the root of a large group of related mtDNAs (the 'R haplogroup'), which occurs today across Eurasia (Supplementary Information section 8). The Y chromosome sequence of the Ust'-Ishim individual is similarly inferred to be ancestral to a group of related Y chromosomes (haplogroup K(xLT)) that occurs across Eurasia today<sup>6</sup> (Supplementary Information section 9). As expected, the number of mutations inferred to have occurred on the

branch leading to the Ust'-Ishim mtDNA is lower than the numbers inferred to have occurred on the branches leading to related present-day mtDNAs (Supplementary Fig. 8.1). Using this observation and nine directly carbon-dated ancient modern human mtDNAs as calibration points<sup>5,7</sup> in a relaxed molecular clock model, we estimate the age of the Ust'-Ishim bone to be  $\sim 49,000$  years BP (95% highest posterior density: 31,000–66,000 years BP), consistent with the radiocarbon date.

In a principal component analysis of the Ust'-Ishim autosomal genome along with genotyping data from 922 present-day individuals from 53 populations<sup>8</sup> (Fig. 2a), the Ust'-Ishim individual clusters with non-Africans rather than Africans. When only non-African populations are analysed (Fig. 2b), the Ust'-Ishim individual falls close to zero on the two first principal component axes, suggesting that it does not share much more ancestry with any particular group of present-day humans. To determine how the Ust'-Ishim genome is related to the genomes of present-day humans, we tested, using *D* statistics<sup>8</sup>, whether it shares more derived alleles with one modern human than with another modern human using pairs of human genomes from different parts of the world (Fig. 3). Based on genotyping data for 87 African and 108 non-African individuals (Supplementary Information section 11), the Ust'-Ishim genome shares more alleles with non-Africans than with sub-Saharan Africans ( $|Z| = 41\text{--}89$ ), consistent with the principal component analysis, mtDNA and Y chromosome results. Thus, the Ust'-Ishim individual represents a population derived from, or related to, the population involved in the dispersal of modern humans out of Africa. Among the non-Africans,





**Figure 2 | Principal Components (PC) analysis exploring the relationship of Ust'-Ishim to present-day humans.** **a**, PC analysis using 922 present-day individuals from 53 populations and the Ust'-Ishim individual. **b**, PC analysis

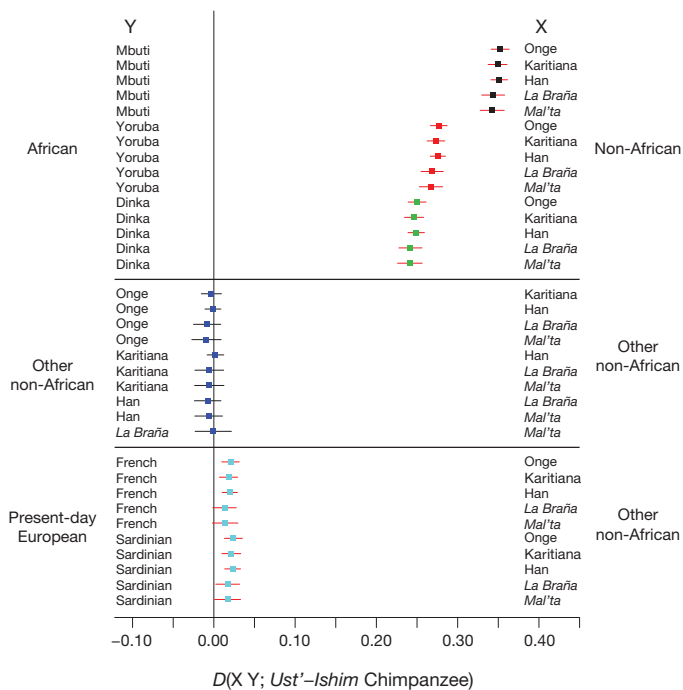
the Ust'-Ishim genome shares more derived alleles with present-day people from East Asia than with present-day Europeans ( $|Z| = 2.1\text{--}6.4$ ). However, when an  $\sim 8,000$ -year-old genome from western Europe (La Braña)<sup>9</sup> or a 24,000-year-old genome from Siberia (Mal'ta 1)<sup>10</sup> were analysed, there is no evidence that the Ust'-Ishim genome shares more derived alleles with present-day East Asians than with these prehistoric individuals ( $|Z| < 2$ ). This suggests that the population to which the Ust'-Ishim individual belonged diverged from the ancestors of present-day

using Eurasian individuals and the Ust'-Ishim individual. The percentages of the total variance explained by each eigenvector are given.

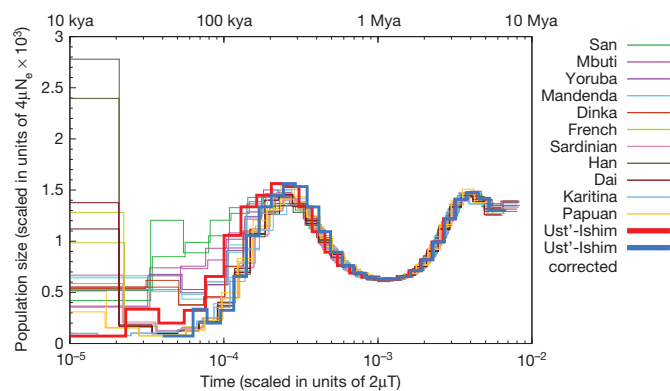
West Eurasian and East Eurasian populations before—or simultaneously with—their divergence from each other. The finding that the Ust'-Ishim individual is equally closely related to present-day Asians and to 8,000- to 24,000-year-old individuals from western Eurasia, but not to present-day Europeans, is compatible with the hypothesis that present-day Europeans derive some of their ancestry from a population that did not participate in the initial dispersals of modern humans into Europe and Asia<sup>11</sup>.

### Mutation rate estimates

The high-quality Ust'-Ishim genome sequence, in combination with its radiocarbon date, allows us to gauge the rate of mutations by estimating the numbers of mutations that are 'missing' in the Ust'-Ishim individual relative to present-day humans. This results in a mutation rate estimate of  $0.44 \times 10^{-9}$  to  $0.63 \times 10^{-9}$  per site per year using the high-coverage genomes of 14 present-day humans. A challenge for inferring the mutation rate in this way is that differences in error rates among genome sequences can confound the inference (see discussion in ref. 12). We therefore developed an alternative approach that leverages the Pairwise Sequentially Markovian Coalescent (PSMC), a method which estimates the distribution of coalescence times between the two chromosomes across a diploid genome to estimate past changes in population size<sup>13</sup>, and which is less influenced by differences in error rates. When the Ust'-Ishim genome along with 25 present-day human genomes are analysed by PSMC, a recent reduction in population size similar to that seen for 11 present-day non-Africans is inferred for the Ust'-Ishim genome. However, the apparent age of this size reduction is more recent than in present-day humans, consistent with the Ust'-Ishim genome being older (Fig. 4). We then compute the number of additional substitutions that are needed to best fit the Ust'-Ishim PSMC curve to those of other non-African genomes. Assuming that this corresponds to the number of mutations that have accumulated over around 45,000 years, we estimate a mutation rate of  $0.43 \times 10^{-9}$  per site per year (95% CI  $0.38 \times 10^{-9}$  to  $0.49 \times 10^{-9}$ ) that is consistent across all non-African genomes regardless of their coverage (Supplementary Information section 14). This overall rate, as well as the relative rates inferred for different mutational classes (transversions, non-CpG transitions, and CpG transitions), is similar to the rate observed for *de novo* estimates from human pedigrees ( $\sim 0.5 \times 10^{-9}$  per site per year<sup>14,15</sup>) and to the direct estimate of branch shortening (Supplementary Information section 10). As discussed elsewhere<sup>14,16,17</sup>, these rates are slower than those estimated using calibrations based on the fossil record and thus suggest older dates for the splits of modern



**Figure 3 | Statistics testing whether the Ust'-Ishim genome shares more derived alleles with one or the other of two modern human genomes (X, Y).** We computed  $D$  statistics of the form  $D(X, Y, \text{Ust'-Ishim, Chimpanzee})$  using a subset of the genome-wide SNP array data from the Affymetrix Human Origins array and restricting the analysis to transversions. Error bars correspond to three standard errors. Red bars indicate that the  $D$  statistic is significantly different from 0 ( $|Z| > 2$ ), such that the Ust'-Ishim genome shares more derived alleles with the genome on the right (X) than the left (Y). Ancient genomes are given in italics.



**Figure 4 | Inferred population size changes over time.** ‘Time’ on the *x* axis refers to the pairwise per-site sequence divergence. If we erroneously assume that Ust'-Ishim lived today, its inferred population size history includes an out-of-Africa-like population bottleneck that is more recent than that seen in present-day non-Africans (red bold curve). By shifting the Ust'-Ishim curve to align with those in present-day non-Africans (blue bold curve), and assuming that the number of mutations necessary to do this corresponds to 45,000 years, we estimate the autosomal mutation rate to be  $0.38 \times 10^{-9}$  to  $0.49 \times 10^{-9}$  per site per year. The times indicated on the top of the figure are based on this mutation rate.

human and archaic populations. We caution, however, that rates may have changed over time and may differ between human populations. However, we expect this mutation rate estimate to apply at least to non-African populations over the past 45,000 years.

We also estimated a phylogeny relating the non-recombining part of the Ust'-Ishim Y chromosome to those of 23 present-day males. Using this phylogeny, we measured the number of ‘missing’ mutations in the Ust'-Ishim Y chromosomal lineage relative to the most closely related present-day Y chromosome analysed. This results in an estimate of the Y chromosome mutation rate of  $0.76 \times 10^{-9}$  per site per year (95% CI  $0.67 \times 10^{-9}$  to  $0.86 \times 10^{-9}$ ) (Supplementary Information section 9), significantly higher than the autosomal mutation rate, consistent with mutation rates in males being higher than in females<sup>18–20</sup>. Finally, using the radiocarbon date of the Ust'-Ishim femur together with the mtDNAs of 311 present-day humans, we estimated the mutation rate of the complete mtDNA to be  $2.53 \times 10^{-8}$  substitutions per site per year (95% highest posterior density:  $1.76 \times 10^{-8}$  to  $3.23 \times 10^{-8}$ ) (Supplementary Information section 8) for mtDNA, in agreement with a previous study<sup>5</sup>.

## Neanderthal admixture

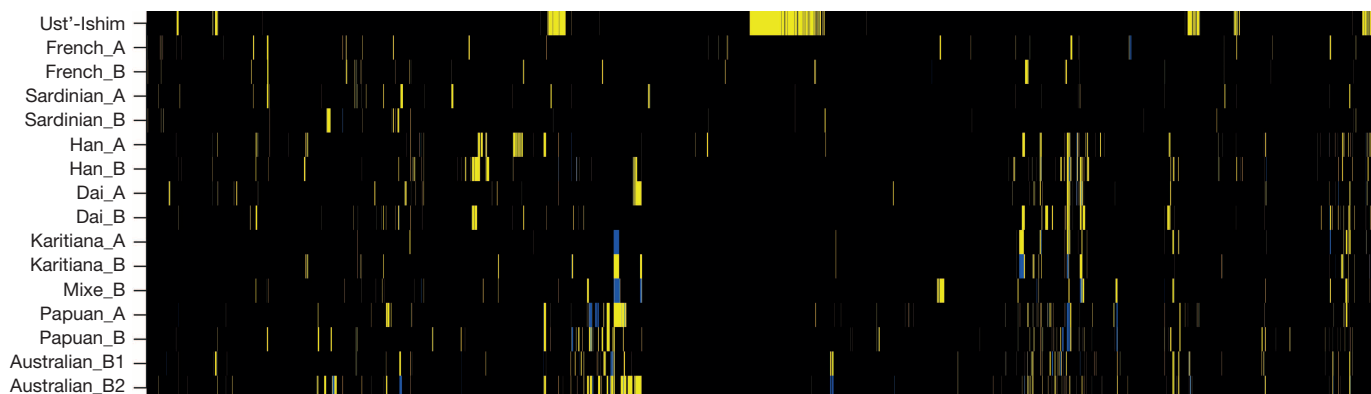
The time of admixture between modern humans and Neanderthals has previously been estimated to 37,000–86,000 years BP based on the size

of the DNA segments contributed by Neanderthals to present-day non-Africans<sup>21</sup>. Thus, the Ust'-Ishim individual could pre-date the Neanderthal admixture. From the extent of sharing of derived alleles between the Neanderthal and the Ust'-Ishim genomes we estimate the proportion of Neanderthal admixture in the Ust'-Ishim individual to be  $2.3 \pm 0.3\%$  (Supplementary Information section 16), similar to present-day east Asians (1.7–2.1%) and present-day Europeans (1.6–1.8%). Thus, admixture with Neanderthals had already occurred by 45,000 years ago. In contrast, we fail to detect any contribution from Denisovans, although such a contribution exists in present-day people not only in Oceania<sup>22,23</sup>, but to a lesser extent also in mainland east Asia<sup>12,24</sup> (Supplementary Information section 17).

The DNA segments contributed by Neanderthals to the Ust'-Ishim individual are expected to be longer than such segments in present-day people as the Ust'-Ishim individual lived closer in time to when the admixture occurred, so there was less time for the segments to be fragmented by recombination. To test if this is indeed the case, we identified putative Neanderthal DNA segments in the Ust'-Ishim and present-day genomes based on derived alleles shared with the Neanderthal genome at positions where Africans are fixed for ancestral alleles. Figure 5 shows that fragments of putative Neanderthal origin in the Ust'-Ishim individual are substantially longer than those in present-day humans. We use the covariance in such derived alleles of putative Neanderthal origin across the Ust'-Ishim genome to infer that mean fragment sizes in the Ust'-Ishim genome are in the order of  $\sim 1.8$ – $4.2$  times longer than in present-day genomes and that the Neanderthal gene flow occurred 232–430 generations before the Ust'-Ishim individual lived (Supplementary Information section 18; Fig. 6). Under the simplifying assumption that the gene flow occurred as a single event, and assuming a generation time of 29 years<sup>16,25</sup>, we estimate that the admixture between the ancestors of the Ust'-Ishim individual and Neanderthals occurred approximately 50,000 to 60,000 years BP, which is close to the time of the major expansion of modern humans out of Africa and the Middle East. However, we also note that the presence of some longer fragments (Fig. 5) may indicate that additional admixture occurred even later. Nevertheless, these results suggest that the bulk of the Neanderthal contribution to present-day people outside Africa does not go back to mixture between Neanderthals and the anatomically modern humans who lived in the Middle East at earlier times; for example, the modern humans whose remains have been found at Skhul and Qafzeh<sup>26,27</sup>.

## An Initial Upper Paleolithic individual?

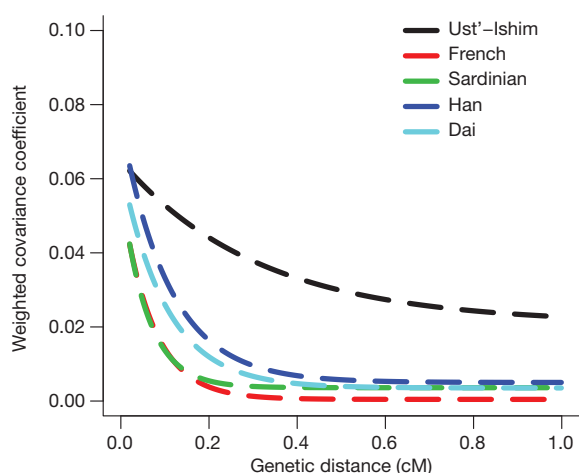
A common model for the modern human colonization of Asia<sup>23,28</sup> assumes that an early coastal migration gave rise to the present-day people of Oceania, while a later more northern migration gave rise to Europeans and mainland Asians. The fact that the 45,000-year-old individual from Siberia is not more closely related to the Onge from the Andaman



**Figure 5 | Regions of Neanderthal ancestry on chromosome 12 in the Ust'-Ishim individual and fifteen present-day non-Africans.** The analysis is based on SNPs where African genomes carry the ancestral allele and the

Neanderthal genome carries the derived allele. Homozygous ancestral alleles are black, heterozygous derived alleles yellow, and homozygous derived alleles blue.





**Figure 6 | Dating the Neanderthal admixture in Ust'-Ishim and present-day non-Africans.** Exponentially fitted curves showing the decay of pairwise covariance for variable positions where Africans carry ancestral alleles and the Neanderthal genome carries derived alleles.

Islands (putative descendants of an early coastal migration) than he is to present-day East Asians or Native Americans (putative descendants of a northern migration) (Fig. 3) shows that at least one other group to which the ancestors of the Ust'-Ishim individual belonged colonized Asia before 45,000 years ago. Interestingly, the Ust'-Ishim individual probably lived during a warm period (Greenland Interstadial 12) that has been proposed to be a time of expansion of modern humans into Europe<sup>29,30</sup>. However, the latter hypothesis is based only on the appearance of the so-called 'Initial Upper Paleolithic' industries (Supplementary Information section 5), and not on the identification of modern human remains<sup>31,32</sup>. It is possible that the Ust'-Ishim individual was associated with the Asian variant of Initial Upper Paleolithic industry, documented at sites such as Kara-Bom in the Altai Mountains at about 47,000 years BP. This individual would then represent an early modern human radiation into Europe and Central Asia that may have failed to leave descendants among present-day populations<sup>29</sup>.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 15 May; accepted 29 August 2014.**

- Trinkaus, E. & Ruff, C. B. Diaphyseal cross-sectional geometry of Near Eastern Middle Paleolithic humans: the femur. *J. Archaeol. Sci.* **26**, 409–424 (1999).
- Brock, F. *et al.* Reliability of nitrogen content (%N) and carbon:nitrogen atomic ratios (C:N) as indicators of collagen preservation suitable for radiocarbon dating. *Radiocarbon* **54**, 879–886 (2012).
- Richards, M. P. & Trinkaus, E. Out of Africa: modern human origins special feature: isotopic evidence for the diets of European Neanderthals and early modern humans. *Proc. Natl Acad. Sci. USA* **106**, 16034–16039 (2009).
- Meyer, M. *et al.* A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222–226 (2012).
- Fu, Q. *et al.* A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr. Biol.* **23**, 553–559 (2013).
- The Y Chromosome Consortium A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res.* **12**, 339–348 (2002).
- Shapiro, B. *et al.* A Bayesian phylogenetic method to estimate unknown sequence ages. *Mol. Biol. Evol.* **28**, 879–887 (2011).
- Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).
- Olalde, I. *et al.* Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature* **507**, 225–228 (2014).
- Raghavan, M. *et al.* Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505**, 87–91 (2014).
- Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409–413 (2014).
- Prüfer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43–49 (2014).

- Li, H. & Durbin, R. Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493–496 (2011).
- Scally, A. & Durbin, R. Revising the human mutation rate: implications for understanding human evolution. *Nature Rev. Genet.* **13**, 745–753 (2012).
- Kong, A. *et al.* Rate of *de novo* mutations and the importance of father's age to disease risk. *Nature* **488**, 471–475 (2012).
- Langergraber, K. E. *et al.* Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution. *Proc. Natl Acad. Sci. USA* **109**, 15716–15721 (2012).
- Prüfer, K. *et al.* The bonobo genome compared with the chimpanzee and human genomes. *Nature* **486**, 527–531 (2012).
- Xue, Y. *et al.* Human Y chromosome base-substitution mutation rate measured by direct sequencing in a deep-rooting pedigree. *Curr. Biol.* **19**, 1453–1457 (2009).
- Kuroki, Y. *et al.* Comparative analysis of chimpanzee and human Y chromosomes unveils complex evolutionary pathway. *Nature Genet.* **38**, 158–167 (2006).
- Wang, J., Fan, H. C., Behr, B. & Quake, S. R. Genome-wide single-cell analysis of recombination activity and *de novo* mutation rates in human sperm. *Cell* **150**, 402–412 (2012).
- Sankararaman, S., Patterson, N., Li, H., Pääbo, S. & Reich, D. The date of interbreeding between Neanderthals and modern humans. *PLoS Genet.* **8**, e1002947 (2012).
- Reich, D. *et al.* Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**, 1053–1060 (2010).
- Reich, D. *et al.* Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania. *Am. J. Hum. Genet.* **89**, 516–528 (2011).
- Skoglund, P. & Jakobsson, M. Archaic human ancestry in East Asia. *Proc. Natl Acad. Sci. USA* **108**, 18301–18306 (2011).
- Fenner, J. N. Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. *Am. J. Phys. Anthropol.* **128**, 415–423 (2005).
- McCown, T. D. & Keith, A. *The Stone Age of Mount Carmel* Vol. 2 (Clarendon, Oxford, 1939).
- Vandermeersch, B. *Les Hommes Fossiles de Qafzeh (Israel)* 319 (Éditions du CNRS, 1981).
- Rasmussen, M. *et al.* An Aboriginal Australian genome reveals separate human dispersals into Asia. *Science* **334**, 94–98 (2011).
- Hublin, J. J. The earliest modern human colonization of Europe. *Proc. Natl Acad. Sci. USA* **109**, 13471–13472 (2012).
- Müller, U. C. *et al.* The role of climate in the spread of modern humans into Europe. *Quat. Sci. Rev.* **30**, 273–279 (2011).
- Goebel, T. A., Derevianko, A. P. & Petrin, V. T. Dating the Middle to Upper Paleolithic transition at Kara-Bom. *Curr. Anthropol.* **34**, 452–458 (1993).
- Kuhn, S. L. & Zwyns, N. Rethinking the initial Upper Paleolithic. *Quat. Int.* <http://dx.doi.org/10.1016/j.quaint.2014.05.040> (2014).
- Bronk Ramsey, C., Scott, M. & van der Plicht, H. Calibration for archaeological and environmental terrestrial samples in the time range 26–50 ka cal BP. *Radiocarbon* **55**, 2021–2027 (2013).
- Reimer, P. J. *et al.* IntCal13 and Marine13 radiocarbon age calibration curves 0–50,000 Years cal BP. *Radiocarbon* **55**, 1869–1887 (2009).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We are grateful to P. Gunz, M. Kircher, A. I. Krivoschapkin, P. Nigst, M. Ongyerth, N. Patterson, G. Renaud, U. Stenzel, M. Stoneking and S. Talamo for valuable input, comments and help; T. Pfisterer and H. Temming for technical assistance. Q.F. is funded in part by the Chinese Academy of Sciences (XDA05130202) and the Ministry of Science and Technology of China (2007FY110200); P.A.K. by Urals Branch, Russian Academy of Sciences (12-C-4-1014) and Y.V.K. by the Russian Foundation for Basic Sciences (12-06-00045); F.J. and M.S. by the National Institutes of Health of the USA (R01-GM40282); P.J. by the NIH (K99-GM104158); and T.F.G.H. by ERC advanced grant 324139. D.R. is a Howard Hughes Medical Institute Investigator and supported by the National Science Foundation (1032255) and the NIH (GM100233). Major funding for this work was provided by the Presidential Innovation Fund of the Max Planck Society.

**Author Contributions** Q.F., S.M.S., A.A.B., Y.V.K., J.K., T.B.V. and S.P. designed the research. A.A.P. and Q.F. performed the experiments; Q.F., H.L., P.M., F.J., P.L.F.J., K.P., C.d.F., M.M., M.L., M.S., D.R., J.K. and S.P. analysed genetic data; K.D. and T.F.G.H. performed <sup>14</sup>C dating; D.C.S.-G. and M.P.R. analysed stable isotope data; N.V.P., P.A.K. and D.I.R. contributed samples and data; S.M.S., A.A.B., N.Z., Y.V.K., S.G.K., J.-J.H. and T.B.V. analysed archaeological and anthropological data; Q.F., J.K., T.B.V. and S.P. wrote and edited the manuscript with input from all authors.

**Author Information** All sequence data have been submitted to the European Nucleotide Archive (ENA) and are available under the following Ust'-Ishim accession number: PRJEB6622. The data from the 25 present-day human genomes are available from (<http://www.simonsonfoundation.org/life-sciences/simons-genome-diversity-project/>) and from (<http://cdna.eva.mpg.de/neandertal/altai/>). Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Q.F. (qiaomei\_fu@eva.mpg.de), D.R. (reich@genetics.med.harvard.edu), J.K. (kelso@eva.mpg.de) or T.B.V. (bence\_viola@eva.mpg.de).

## METHODS

All sequencing was performed on the Illumina HiSeq 2000 and base-calling was carried out using Ibis 1.1.6.9 (ref. 35). Reads were merged and remaining adaptor sequences trimmed before being aligned to the Human reference genome (GRCh37/1000 Genomes) using BWA (version 0.5.10)<sup>36</sup>. GATK version 1.3 (v1.3-14-g348f2b) was used to produce genotype calls for each site. We excluded from analysis tandem repeats and regions of the genome that are not unique. We considered only

genomic regions that fall within the 95% coverage distribution (Supplementary Information section 7) and where at least 99% of overlapping 35mers covering a position map uniquely, allowing one mismatch.

35. Kircher, M., Stenzel, U. & Kelso, J. Improved base calling for the Illumina Genome Analyzer using machine learning strategies. *Genome Biol.* **10**, R83 (2009).
36. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).



# Pulmonary macrophage transplantation therapy

Takuji Suzuki<sup>1</sup>, Paritha Arumugam<sup>2</sup>, Takuro Sakagami<sup>1</sup>, Nico Lachmann<sup>3</sup>, Claudia Chalk<sup>1</sup>, Anthony Salles<sup>1</sup>, Shuichi Abe<sup>1</sup>, Cole Trapnell<sup>4,5</sup>, Brenna Carey<sup>1</sup>, Thomas Moritz<sup>3</sup>, Punam Malik<sup>2</sup>, Carolyn Lutzko<sup>2</sup>, Robert E. Wood<sup>6</sup> & Bruce C. Trapnell<sup>1,6,7</sup>

**Bone-marrow transplantation is an effective cell therapy but requires myeloablation, which increases infection risk and mortality. Recent lineage-tracing studies documenting that resident macrophage populations self-maintain independently of haematological progenitors prompted us to consider organ-targeted, cell-specific therapy. Here, using granulocyte-macrophage colony-stimulating factor (GM-CSF) receptor- $\beta$ -deficient (*Csf2rb*<sup>-/-</sup>) mice that develop a myeloid cell disorder identical to hereditary pulmonary alveolar proteinosis (hPAP) in children with *CSF2RA* or *CSF2RB* mutations, we show that pulmonary macrophage transplantation (PMT) of either wild-type or *Csf2rb*-gene-corrected macrophages without myeloablation was safe and well-tolerated and that one administration corrected the lung disease, secondary systemic manifestations and normalized disease-related biomarkers, and prevented disease-specific mortality. PMT-derived alveolar macrophages persisted for at least one year as did therapeutic effects. Our findings identify mechanisms regulating alveolar macrophage population size in health and disease, indicate that GM-CSF is required for phenotypic determination of alveolar macrophages, and support translation of PMT as the first specific therapy for children with hPAP.**

Mutations in *CSF2RA* or *CSF2RB*, encoding GM-CSF receptor  $\alpha$  or  $\beta$ , respectively, cause hPAP by impairing GM-CSF-dependent surfactant clearance by alveolar macrophages, resulting in progressive surfactant accumulation in alveoli and hypoxaemic respiratory failure<sup>1–5</sup>. Surfactant normally comprises a thin phospholipid/protein layer reducing tension on the alveolar surface<sup>6</sup> that is maintained by balanced secretion by alveolar type II epithelial cells and clearance by these cells and alveolar macrophages<sup>7,8</sup>. PAP also occurs in people with GM-CSF autoantibodies (~85–90% of all patients with PAP)<sup>9,10</sup> and mice with disruption of the GM-CSF gene *Csf2* (refs 11, 12) or the GM-CSF receptor  $\beta$ -subunit gene *Csf2rb* (refs 13, 14) (*Csf2*<sup>-/-</sup> or *Csf2rb*<sup>-/-</sup> mice, respectively). Characteristics of PAP caused by disruption of GM-CSF signalling include typical lung histopathology (well preserved alveoli filled with surfactant and ‘foamy’ macrophages staining positive with periodic acid-Schiff (PAS) or oil red O); turbid, ‘milky’ appearing bronchoalveolar lavage (BAL) caused by accumulated surfactant and cell debris; and a disease-specific pattern of biomarkers (increased GM-CSF (hPAP), M-CSF/CSF1 and MCP-1 in BAL fluid, and reduced mRNA for *PU.1*, *PPARG* and *ABCG1* in alveolar macrophages)<sup>1–5,15–20</sup>.

Currently, no pharmacological therapy of hPAP exists and surfactant must be removed by whole-lung lavage, an inefficient, invasive procedure to physically remove excess surfactant<sup>2–4</sup>. In *Csf2rb*<sup>-/-</sup> mice, PAP was corrected by bone marrow transplantation (BMT) of wild-type (WT)<sup>21</sup> or *Csf2rb*-gene-corrected *Csf2rb*<sup>-/-</sup> haematopoietic stem/progenitor cells (HSPCs)<sup>22</sup>. However, in humans this approach resulted in death from infection before engraftment<sup>2</sup>, probably as a result of required myeloablation/immunosuppressive therapy. Since pulmonary GM-CSF is increased in hPAP<sup>1–5</sup> we hypothesized that macrophages administered directly into the lungs (pulmonary macrophage transplantation or PMT) without myeloablation would engraft and reverse the manifestations of hPAP.

We first validated *Csf2rb*<sup>-/-</sup> mice as a model of human hPAP by demonstrating that they had the same clinical, physiological, histopathological

and biochemical abnormalities, disease biomarkers and natural history (Fig. 1 and Extended Data Fig. 1) as children with hPAP<sup>3</sup>.

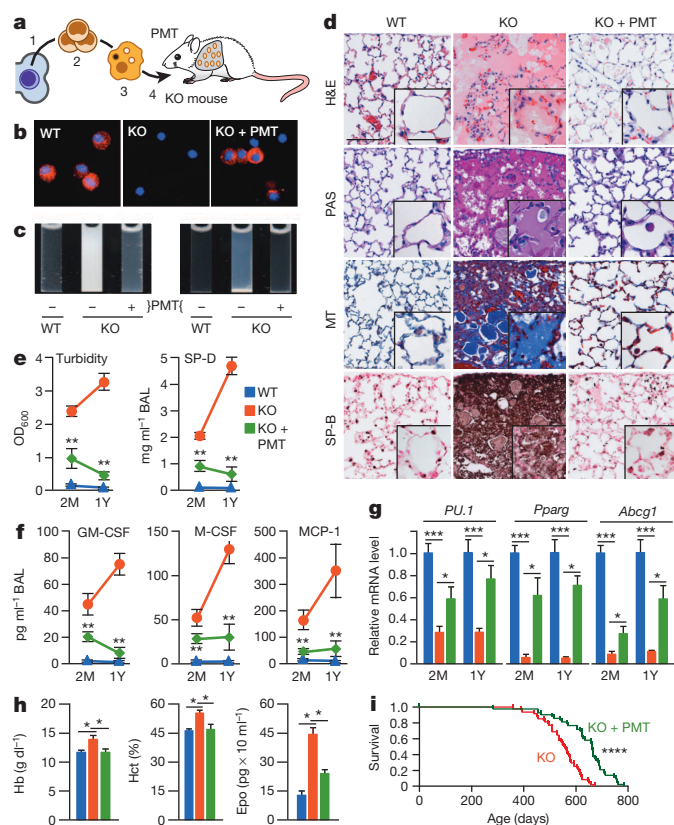
## Characterization of macrophages before PMT

Bone-marrow-derived macrophages (BMDMs) from WT mice had morphology and phenotypic markers (F4/80<sup>+</sup>, CD11b<sup>Hi</sup>, CD11c<sup>+</sup>, CD14<sup>+</sup>, CD16/32<sup>+</sup>, CD64<sup>+</sup>, CD68<sup>+</sup>, CD115<sup>+</sup>, CD131<sup>+</sup>, SiglecF<sup>Low</sup>, MerTK<sup>+</sup>, MHC class II<sup>+</sup>, Ly6G<sup>-</sup>, CD3<sup>-</sup>, CD19<sup>-</sup>) of macrophages (Extended Data Fig. 2a–c) and contained <0.0125% lineage negative (Lin<sup>-</sup>) Sca1<sup>+</sup>c-Kit<sup>+</sup> (LSK) cells. Clonogenic analysis indicated <0.005% colony-forming units-granulocyte, monocyte/macrophage (CFU-GM) and no burst-forming units-erythrocyte (BFU-E) or colony-forming units-granulocyte, erythrocyte, monocyte/macrophage, megakaryocyte (CFU-GEMM) progenitors (Extended Data Fig. 2d, e). Functional evaluation<sup>23</sup> showed that the BMDMs could clear surfactant (Extended Data Fig. 2f, g). These results demonstrated that the cells used for PMT were highly purified, mature macrophages capable of surfactant clearance.

## Efficacy of PMT of WT macrophages

To determine the therapeutic potential of PMT, *Csf2rb*<sup>-/-</sup> mice received WT (*Csf2rb*<sup>+/+</sup>) BMDMs by PMT once (Fig. 1a). One year later, PMT-derived CD131<sup>+</sup> BAL cells were present (Fig. 1b), alveolar macrophages expressed *Csf2rb* (Extended Data Fig. 3a), and BAL was markedly improved with respect to opacification (Fig. 1c), sediment (Fig. 1c) and microscopic cytopathology (Extended Data Fig. 3b). Importantly, PMT nearly completely resolved the abnormal pulmonary histopathology (Fig. 1d and Extended Data Fig. 3c). Measurement of BAL turbidity and surfactant protein-D (SP-D) content (Fig. 1e), which reflect the extent of surfactant accumulation across the entire lung surface, confirmed the improvement in hPAP. BAL fluid biomarkers of hPAP were also improved (Fig. 1f). The effects of PMT were evident early, as demonstrated by detection of CD131<sup>+</sup> alveolar macrophages with *Csf2rb* mRNA and protein (not

<sup>1</sup>Division of Pulmonary Biology, Perinatal Institute, Cincinnati Children's Hospital Medical Center, 3333 Burnet Avenue, Cincinnati, Ohio 45229, USA. <sup>2</sup>Division of Experimental Hematology, Cincinnati Children's Hospital Medical Center, 3333 Burnet Avenue, Cincinnati, Ohio 45229, USA. <sup>3</sup>RG Reprogramming and Gene Therapy, Institute of Experimental Hematology, Hannover Medical School, Carl Neuberg-Str. 1, 30625 Hannover, Germany. <sup>4</sup>Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, Massachusetts 02138, USA. <sup>5</sup>Broad Institute of Massachusetts Institute of Technology and Harvard University, Cambridge, Massachusetts 02138, USA. <sup>6</sup>Division of Pulmonary Medicine, Cincinnati Children's Hospital Medical Center, 3333 Burnet Avenue, Cincinnati, Ohio 45229, USA. <sup>7</sup>Division of Pulmonary, Critical Care, and Sleep Medicine, University of Cincinnati Medical Center, 3333 Burnet Avenue, Cincinnati, Ohio 45229, USA.



**Figure 1 | Therapeutic efficacy of PMT in *Csf2rb*<sup>-/-</sup> mice.** **a**, Schematic of the method used. WT HSPCs (1) were isolated, expanded (2), differentiated into macrophages (3), and administered by endotracheal instillation into 2-month-old *Csf2rb*<sup>-/-</sup> (KO) mice (4) and evaluated after 2 months (2M) (e–g) or one year (1Y) (b–h) with age-matched, untreated WT or *Csf2rb*<sup>-/-</sup> mice (KO+PMT, WT or KO, respectively). **b**, CD131-immunostained BAL cells. **c**, Appearance of BAL fluid (left) or sediment (right). **d**, Lung histology after staining with haematoxylin and eosin (H&E), PAS, Masson's trichrome (MT), or surfactant protein B (SP-B). Scale bar, 100  $\mu$ m; inset, 50  $\mu$ m. **e**, BAL turbidity and SP-D concentration. **f**, BAL biomarkers. **g**, Alveolar macrophage biomarkers. **h**, Effects of PMT on blood haemoglobin (Hb), haematocrit (Hct) and serum erythropoietin (Epo). **i**, Kaplan–Meier analysis of PMT-treated ( $n = 43$ ) and untreated *Csf2rb*<sup>-/-</sup> mice ( $n = 48$ ). Images are representative of 6 mice per group (b–d). Numeric data are mean  $\pm$  s.e.m. of 7 (2M) or 6 (1Y) mice per group. \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ , \*\*\*\* $P < 0.0001$ .

shown), reduced BAL opacification and cytopathology (not shown), BAL turbidity (Fig. 1e), SP-D (Fig. 1e) and BAL fluid biomarkers (Fig. 1f) 2 months after PMT, and reduced lung histopathology 4 months after PMT (not shown). In contrast, PMT of *Csf2rb*<sup>-/-</sup> BMDMs had no effect on BAL turbidity, SP-D content, or BAL fluid biomarkers (not shown), demonstrating that GM-CSF receptors on transplanted macrophages are important for the therapeutic effects.

To evaluate the effects of PMT on the alveolar macrophage population, we measured cellular biomarkers after PMT. Results showed that alveolar macrophages from PMT-treated *Csf2rb*<sup>-/-</sup> mice had increased mRNA for *PU.1*, *Pparg* and *Abcg1*, improvement was significant by 2 months, and the effects persisted 1 year after PMT (Fig. 1g).

Since *Csf2rb*<sup>-/-</sup> mice develop polycythaemia, a secondary consequence of hypoxaemia in chronic lung diseases<sup>24</sup>, we evaluated the effects of PMT on this systemic clinical manifestation. Notably, PMT corrected polycythaemia in *Csf2rb*<sup>-/-</sup> mice (Fig. 1h).

Finally, we evaluated the effects of PMT on hPAP-associated mortality by comparing the survival of PMT-treated and untreated *Csf2rb*<sup>-/-</sup> mice. PMT increased the lifespan of *Csf2rb*<sup>-/-</sup> mice by 107 days, from 555 (median; interquartile range 507–592) days to 662 (604–692) days (Fig. 1i). In separate studies of treated *Csf2rb*<sup>-/-</sup> mice surviving to 617

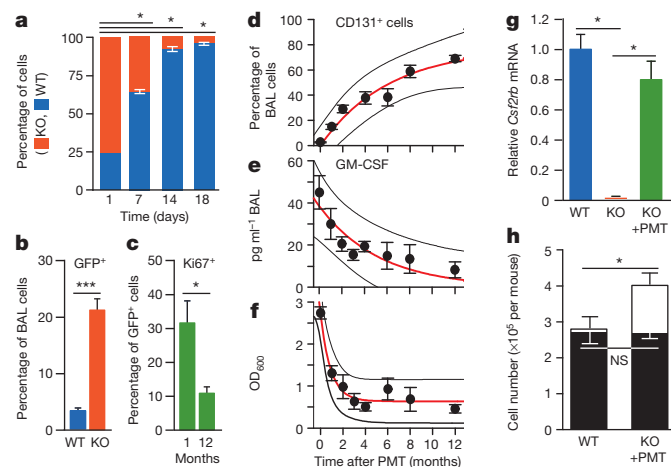
(604–631) days (561 (548–575) days after PMT of WT BMDMs), CD131<sup>+</sup> alveolar macrophages were still present and BAL turbidity remained low compared to untreated *Csf2rb*<sup>-/-</sup> mice that survived to 631 (631–631) days (optical density at 600 nm (OD<sub>600</sub>) = 0.75  $\pm$  0.17 versus 2.63  $\pm$  0.44;  $n = 8, 4$ , respectively;  $P < 0.001$ ). However, such long-term evaluation of laboratory abnormalities is obfuscated by reduced survival of untreated *Csf2rb*<sup>-/-</sup> mice.

These results demonstrate that PMT had a highly efficacious and durable therapeutic effect on the primary pulmonary and secondary systemic manifestations of hPAP in *Csf2rb*<sup>-/-</sup> mice.

## Macrophage engraftment efficiency

We next evaluated the effects of cell dose (0.5, 1, 2 and 4 million) and repeated administration (one versus four monthly transplantations) on PMT efficacy (Extended Data Tables 2 and 3, respectively). Neither treatment significantly affected efficacy in the range evaluated, and one dose of 2 million cells was used for PMT in the remaining studies.

To determine whether WT macrophages had a survival advantage over *Csf2rb*<sup>-/-</sup> macrophages, we measured GM-CSF bioactivity in BAL fluid and found that it was detectable in *Csf2rb*<sup>-/-</sup> but not WT mice (Extended Data Fig. 1h). WT macrophages had increased survival/proliferation compared to *Csf2rb*<sup>-/-</sup> macrophages *in vitro* (Fig. 2a) and accumulated to greater numbers after PMT in *Csf2rb*<sup>-/-</sup> mice than in WT mice (Fig. 2b and Extended Data Fig. 3d). PMT of WT Lys-M<sup>GFP</sup> knock-in mouse<sup>25</sup> BMDMs into *Csf2rb*<sup>-/-</sup> mice followed by Ki67 immunostaining revealed that PMT-derived cells replicated *in vivo* (Extended Data Fig. 3e–g). The percentage of Ki67<sup>+</sup> PMT-derived alveolar macrophages was 32.2  $\pm$  6.05% 1 month after PMT and declined to 11.29  $\pm$  2.2% by 1 year (Fig. 2c) similar to baseline Ki67<sup>+</sup> immunostaining of alveolar macrophages in age-matched, normal WT mice (Extended Data Fig. 3f). To define this survival advantage further, we evaluated the engraftment kinetics after one PMT of WT BMDMs in *Csf2rb*<sup>-/-</sup> mice. CD131<sup>+</sup> cells increased steadily from zero to 69.0  $\pm$  2.5% of BAL cells (Fig. 2d) synchronous with a smooth decline in pulmonary GM-CSF to



**Figure 2 | Pharmacokinetics and pharmacodynamics of PMT in *Csf2rb*<sup>-/-</sup> mice.** **a**, Competitive proliferation of WT and *Csf2rb*<sup>-/-</sup> BMDMs co-cultured with GM-CSF and M-CSF ( $n = 3$  plates per point). **b**, Quantification of GFP<sup>+</sup> BAL cells 2 months after PMT of Lys-M<sup>GFP</sup> BMDMs into WT ( $n = 3$ ) or *Csf2rb*<sup>-/-</sup> mice ( $n = 6$ ) mice. **c**, Quantification of Ki67<sup>+</sup> Lys-M<sup>GFP</sup> cells in *Csf2rb*<sup>-/-</sup> mice ( $n = 3$ ) 1 or 12 months after PMT. **d–f**, *Csf2rb*<sup>-/-</sup> mice received PMT of WT BMDMs and were evaluated at the indicated times to quantify CD131<sup>+</sup> BAL cells (**d**), BAL GM-CSF concentration (**e**) and BAL turbidity (**f**). Exponential regression ( $\pm$  prediction bands),  $R^2 = 0.943$  (**d**),  $R^2 = 0.819$  (**e**),  $R^2 = 0.958$  (**f**). Data are mean  $\pm$  s.e.m. for 3–7 mice per group. **g**, *Csf2rb* mRNA in BAL cells from *Csf2rb*<sup>-/-</sup> mice 1 year after PMT, or untreated, age-matched control mice ( $n = 6$ ). **h**, Number of BAL cells (open bars) or CD131<sup>+</sup> alveolar macrophages (filled bars) in *Csf2rb*<sup>-/-</sup> mice 1 year after PMT ( $n = 5$ ) or untreated WT mice ( $n = 10$ ). Data are mean  $\pm$  s.e.m. \* $P < 0.05$ , \*\*\* $P < 0.001$ ; NS, not significant.

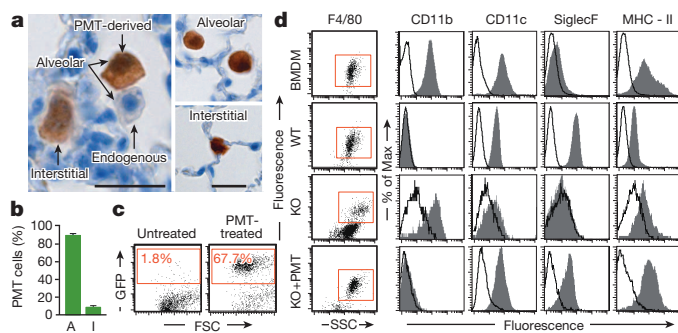


near normal (Fig. 2e). Similarly, BAL turbidity declined with the increase in CD131<sup>+</sup> alveolar macrophages (Fig. 2f). One year after PMT, CD131<sup>+</sup> cells were present (Fig. 1b), CD131 protein (encoded by *Csf2rb*) was detectable in alveolar macrophages (Extended Data Fig. 3a), and *Csf2rb* mRNA in BAL cells from PMT-treated *Csf2rb*<sup>-/-</sup> mice was only slightly less than in WT and undetectable in untreated *Csf2rb*<sup>-/-</sup> BAL cells (Fig. 2g). Importantly, numbers of CD131<sup>+</sup> alveolar macrophages in PMT-treated *Csf2rb*<sup>-/-</sup> and untreated WT mice were similar 1 year after PMT (Fig. 2h). These results demonstrate that WT macrophages had a selective survival advantage over *Csf2rb*<sup>-/-</sup> macrophages and that after PMT into *Csf2rb*<sup>-/-</sup> mice, they proliferated *in vivo* at a rate that slowed over time synchronous with reduction in pulmonary GM-CSF, replaced dysfunctional *Csf2rb*<sup>-/-</sup> alveolar macrophages, and resulted in numbers of CD131<sup>+</sup>, GM-CSF-responsive alveolar macrophages similar to WT mice.

### Macrophage characterization after PMT

The fate of macrophages after PMT was evaluated to determine their spatial distribution, phenotype and gene expression profile. Intra-pulmonary localization was evaluated 1 year after PMT of WT Lys-M<sup>GFP</sup> BMDMs by fluorescence microscopy to identify CD68<sup>+</sup> GFP<sup>+</sup> (that is, PMT-derived) macrophages, which revealed that 88.9 ± 0.87% were intra-alveolar and 11.1 ± 0.87% were interstitial (Extended Data Fig. 3h). GFP immunohistochemical staining was done to eliminate potential interference from autofluorescence and confirmed these results; 90.5 ± 1.1% PMT-derived macrophages were intra-alveolar and 9.4 ± 1.1% were interstitial (Fig. 3a, b and Extended Data Fig. 3i). Localization was done in similarly treated mice by flow cytometry to detect GFP<sup>+</sup> cells 2 months (not shown) or 1 year after PMT (Fig. 3c and Extended Data Fig. 4a, b) and by PCR amplification of Lys-M<sup>GFP</sup> transgene-specific DNA (Extended Data Fig. 4c), all of which showed that PMT-derived cells were present in the lungs but not detected in blood, bone marrow, or spleen. One year after PMT of CD45.1<sup>+</sup> WT BMDMs into CD45.2<sup>+</sup> *Csf2rb*<sup>-/-</sup> mice, flow cytometric detection of CD45.1<sup>+</sup> cells confirmed these findings (Extended Data Fig. 4e–g). Results show that the transplanted macrophages remained in the lungs, primarily within the intra-alveolar space.

The effects of the lung environment on the phenotype of transplanted macrophages were evaluated by measuring cell-surface markers. One year after PMT of WT Lys-M<sup>GFP</sup> BMDMs into *Csf2rb*<sup>-/-</sup> mice, PMT-derived alveolar macrophages comprised 68.7 ± 6.5% of BAL cells and had converted from CD11b<sup>Hi</sup> SiglecF<sup>Low</sup> to CD11b<sup>Low</sup> SiglecF<sup>Hi</sup>, similar to the phenotype of WT alveolar macrophages and different from *Csf2rb*<sup>-/-</sup> mice at the point of PMT (CD11b<sup>Hi</sup> SiglecF<sup>Low</sup>) (Fig. 3c, d). Similarly, one year after PMT of WT CD45.1<sup>+</sup> BMDMs into CD45.2<sup>+</sup> *Csf2rb*<sup>-/-</sup> mice, CD45.1<sup>+</sup> alveolar macrophages comprised 63.6 ± 12.1%



**Figure 3 | Localization and phenotype of transplanted macrophages.** Lys-M<sup>GFP</sup> BMDMs were transplanted into *Csf2rb*<sup>-/-</sup> mice and evaluated after 1 year. **a**, Immunostained lung showing GFP<sup>+</sup> cells. Scale bars, left, 200 μm; right, 20 μm. **b**, Localization of GFP<sup>+</sup> macrophages to intra-alveolar (A) and interstitial (I) spaces (*n* = 6). **c**, GFP<sup>+</sup> BAL cells identified by flow cytometry. **d**, Phenotypic analysis of F4/80<sup>+</sup> BMDMs before PMT, and alveolar macrophages from PMT-treated *Csf2rb*<sup>-/-</sup> mice, or untreated, age-matched WT or *Csf2rb*<sup>-/-</sup> mice (*n* = 6 per group). Data are mean ± s.e.m.

of BAL cells and had undergone the same phenotypic conversion (Extended Data Fig. 4h).

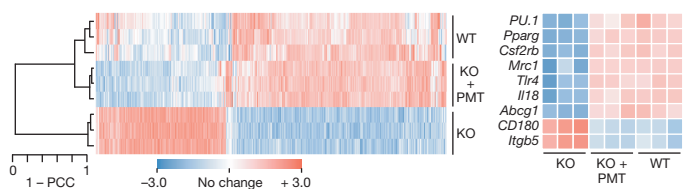
To determine the effects on gene expression, we performed genome-wide expression profiling on alveolar macrophages from *Csf2rb*<sup>-/-</sup> mice 1 year after PMT of WT BMDMs and compared to results for untreated, age-matched WT or *Csf2rb*<sup>-/-</sup> mice. Unsupervised analysis indicated marked co-clustering between PMT-treated *Csf2rb*<sup>-/-</sup> and WT mice while *Csf2rb*<sup>-/-</sup> mice clustered separately (Fig. 4). Expression of genes regulated by GM-CSF was reduced in *Csf2rb*<sup>-/-</sup> mice and restored by PMT (Fig. 4 and Extended Data Fig. 5a). Of 776 genes for which expression was disrupted in *Csf2rb*<sup>-/-</sup> mice, PMT normalized expression of 600 including 80% of genes upregulated and 76% of genes downregulated in *Csf2rb*<sup>-/-</sup> compared to WT mice (Extended Data Fig. 5b). Supervised Gene Ontology (GO) and detailed KEGG pathway analysis revealed that genes in multiple pathways involved in lipid metabolism, cellular proliferation, apoptosis and host defence were coordinately down-regulated in *Csf2rb*<sup>-/-</sup> mice, and normalized by PMT (Extended Data Fig. 5c, d). Results for multiple genes important in lipid metabolism (*Abcg1*, *Nr1h3*, *Olr1*, *Lepr*, *Fabp1*, *Lipf*, *Abca1*, *Apoe*, *Apoc2*, *Pla2g7*) were validated using separate samples (Extended Data Fig. 5e).

### Efficacy of gene therapy by PMT

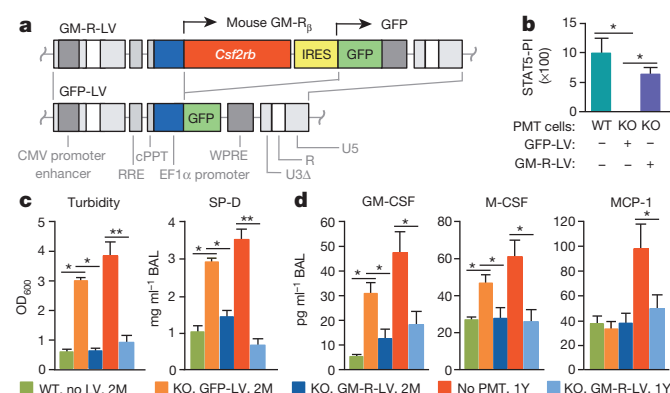
Since PMT in humans would probably employ autologous, gene-corrected HSPC-derived macrophages, we evaluated PMT of *Csf2rb*<sup>-/-</sup> macrophages derived from LSK cells after lentiviral vector (LV)-mediated *Csf2rb* cDNA expression (Fig. 5a). *Csf2rb* gene-corrected (GM-R-LV-transduced) or sham-treated (GFP-LV-transduced) *Csf2rb*<sup>-/-</sup> and non-transduced WT LSK-derived cells all had macrophage morphology, expressed CD68 (Extended Data Fig. 6a) and were F4/80<sup>+</sup> CD11b<sup>Hi</sup> CD11c<sup>+</sup> (not shown). In contrast, only WT and GM-R-LV-transduced *Csf2rb*<sup>-/-</sup> cells were CD131<sup>+</sup> and only lentiviral-vector-transduced cells were GFP<sup>+</sup> (Extended Data Fig. 6a). GM-R-LV restored GM-CSF signalling in *Csf2rb*<sup>-/-</sup> macrophages (Fig. 5b and Extended Data Fig. 6b). Two months after PMT into *Csf2rb*<sup>-/-</sup> mice, GM-CSF receptor-β was detected on alveolar macrophages only from mice receiving gene-corrected *Csf2rb*<sup>-/-</sup> or WT macrophages (Extended Data Fig. 6c). The efficacy using gene-corrected *Csf2rb*<sup>-/-</sup> or WT cells was equivalent as demonstrated by a similar degree of improvement in BAL appearance (Extended Data Fig. 6d), BAL turbidity, SP-D and biomarkers of hPAP (Fig. 5c, d). Furthermore, gene-corrected BMDMs localized to the lung (Extended Data Fig. 4d and Fig. 6e) and underwent phenotypic conversion to CD11b<sup>Low</sup> (Extended Data Fig. 6f). The long-term efficacy of gene-corrected macrophages 1 year after PMT was demonstrated by marked reduction in BAL turbidity, SP-D and BAL fluid biomarkers of hPAP (Fig. 5c, d). These results demonstrate that PMT of gene-corrected macrophages had a therapeutic effect on hPAP in *Csf2rb*<sup>-/-</sup> mice equivalent to that of WT macrophages and was durable, lasting at least 1 year.

### Safety of PMT therapy in *Csf2rb*<sup>-/-</sup> mice

PMT was well tolerated and without adverse effects. One year after PMT, there were no haematological abnormalities (Extended Data Table 4), cellular inflammation or pulmonary fibrosis in mice receiving PMT of WT (Fig. 1d) or gene-corrected macrophages (not shown). *Csf2rb*<sup>-/-</sup>



**Figure 4 | Microarray analysis of alveolar macrophages 1 year after PMT.** Unsupervised hierarchical clustering dendrogram and heat map of selected GM-CSF-regulated genes in PMT-treated *Csf2rb*<sup>-/-</sup> mice or untreated, age-matched WT or *Csf2rb*<sup>-/-</sup> mice (3 per group). Pearson correlation coefficient (PCC).



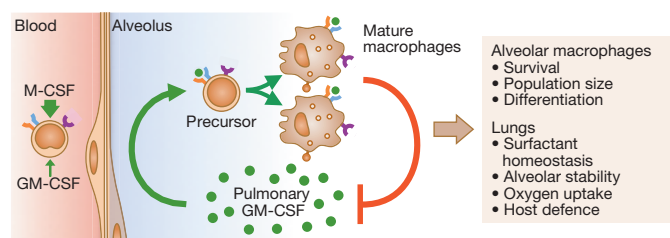
**Figure 5 | Effects of PMT of gene-corrected macrophages on hPAP severity and biomarkers.** *Csf2rb*<sup>-/-</sup> mice received PMT of non-transduced WT or lentiviral-vector-transduced *Csf2rb*<sup>-/-</sup> macrophages and were evaluated after 2 months (2M) or 1 year (1Y) (with untreated, age-matched *Csf2rb*<sup>-/-</sup> mice). The key indicates PMT cells used, previous lentiviral vector treatment, and time after PMT analysis was performed. **a**, Lentiviral vector schematics. **b**, GM-CSF signalling measured by the STAT5 phosphorylation index (STAT5-PI) in the indicated cells before PMT. **c**, BAL turbidity and SP-D concentration. **d**, BAL biomarkers. Mean  $\pm$  s.e.m. of  $n = 3$  (**b**) or 5–10 (**c**, **d**) mice per group. \* $P < 0.05$ , \*\* $P < 0.01$ .

mice had trivial elevations of IL-6 and TNF- $\alpha$  in BAL that were reduced by PMT of WT macrophages (Extended Data Table 4). These data identify no safety concerns for PMT therapy of hPAP in *Csf2rb*<sup>-/-</sup> mice.

## Discussion

Multiple lines of evidence indicate that the high efficacy of PMT therapy of hPAP in *Csf2rb*<sup>-/-</sup> mice was attributable to a selective survival advantage conferred by increased pulmonary GM-CSF to alveolar macrophages bearing functional GM-CSF receptors. However, pulmonary surfactant remained slightly increased 1 year after a single PMT. This could be because the treatment time was too short or exceeded the durability of the clinical benefit, or due to the continued presence of *Csf2rb*<sup>-/-</sup> alveolar macrophages despite engraftment of GM-CSF-responsive macrophages. The latter is likely due to ongoing *Csf2rb*<sup>-/-</sup> myelopoiesis, pulmonary recruitment of monocytes and local proliferation, and GM-CSF-independent survival as occurs in untreated *Csf2rb*<sup>-/-</sup> mice. *Csf2rb*<sup>-/-</sup> macrophages may provide a 'protected intracellular niche' for surfactant accumulation since without GM-CSF, alveolar macrophages internalize but cannot clear surfactant<sup>26</sup>. Another factor may be reduction of the survival advantage over time, that is, reduced pulmonary GM-CSF (driving WT cell proliferation) and reduced surfactant burden (driving surfactant-engorgement-related *Csf2rb*<sup>-/-</sup> cell death). Notwithstanding these points, a single PMT of GM-CSF responsive cells cleared ~90% of the abnormal surfactant accumulation for at least one year.

The feasibility of translating PMT therapy to humans with hPAP is supported by the safety and efficacy of PMT in *Csf2rb*<sup>-/-</sup> mice and the striking similarity of hPAP in mice and humans. Macrophages could be delivered by bronchoscopic instillation without endotracheal intubation, general anaesthesia, or mechanical ventilation, which are required for whole lung lavage and increase risk. Preparative myeloablation would



**Figure 6 | Proposed homeostatic reciprocal feedback mechanism by which pulmonary GM-CSF regulates alveolar macrophage population size in vivo.**

be unnecessary and use of autologous, gene-corrected cells would eliminate the need for immunosuppression, which are required for BMT. PMT may also be possible with gene-corrected, inducible pluripotent-cell-derived macrophages recently prepared from children with hPAP<sup>23,27</sup>. However, formal preclinical toxicology studies related to PMT and to gene transfer will be needed before this approach can be tested in humans. Since pulmonary GM-CSF is critical to lung host defence and clearance of a broad range of microorganisms<sup>28</sup>, PMT may also be useful in treating serious lung infections. Indeed, pulmonary administration of macrophages constitutively expressing IFN- $\gamma$  improved host defence in SCID mice<sup>29</sup>. In such applications, inhaled GM-CSF could be used to promote survival of transplanted macrophages<sup>30</sup>.

Identification of a homeostatic mechanism by which pulmonary GM-CSF regulates alveolar macrophage population size (Fig. 6) was an unexpected but important finding. Its existence is supported by recent fate-mapping studies indicating that tissue-resident alveolar macrophages derive before birth and self-maintain by local replication independent of circulating monocytes at steady state<sup>30–33</sup>.

The concept of alveolar macrophages (and other tissue-resident macrophages) as short-lived, terminally differentiated, non-dividing representatives of a unified mononuclear phagocyte system replenished via monocyte intermediates has evolved considerably since its inception<sup>34</sup>. Alveolar macrophage half-life was initially estimated at 2 weeks based on studies of repopulation after lethal irradiation and allogeneic BMT<sup>35</sup>. Improved detection methods using GFP<sup>+</sup> cells increased the estimate to 30 days<sup>36</sup>. Shielding the thorax during irradiation increased it further to 8 months<sup>37</sup>. Our data, obtained without irradiation or myeloablation, show that macrophages transplanted directly into the respiratory tract persisted for one-and-a-half years. A caveat of such estimates is their inability to discern if persistence is due to prolonged survival or replication.

Normally, alveolar macrophages are phenotypically CD11b<sup>Low</sup>SiglecF<sup>Hi</sup> while other macrophage populations are CD11b<sup>Hi</sup>SiglecF<sup>Low</sup>. Surprisingly, WT BMDMs cultured in GM-CSF and M-CSF were CD11b<sup>Hi</sup>SiglecF<sup>Low</sup> *in vitro* but converted to CD11b<sup>Low</sup>SiglecF<sup>Hi</sup> after PMT. In contrast, BMDMs instilled in the peritoneum adopt the CD11b<sup>Hi</sup> phenotype of peritoneal macrophages<sup>38</sup>. These changes agree with gene-expression profiling studies<sup>39</sup> and indicate that local microenvironments provide critical 'phenotypically instructive' cues that direct development of tissue-resident macrophage populations. Our results show for alveolar macrophages that GM-CSF provides one such phenotypic cue while the lung environment provides another critical, albeit unidentified, cue.

The limitations of our study include the fact that it did not establish a minimum effective dose, a maximum tolerated dose, or a significant dose–response relationship. BMDMs were capable of clearing surfactant before transplantation but results did not determine whether 'lung-conditioning' further increased their clearance capacity. While the macrophages used for PMT contained very few progenitors, it is theoretically possible that clonal expansion of a progenitor subpopulation may have contributed to therapeutic efficacy and, if so, potential clonal shrinkage may have contributed to loss of benefit at later times. Thus, additional studies are needed to further confirm the identity of effector cells and precise pharmacokinetics and durability of the therapeutic benefit.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 12 February; accepted 1 September 2014.

Published online 1 October 2014.

1. Suzuki, T. et al. Familial pulmonary alveolar proteinosis caused by mutations in CSF2RA. *J. Exp. Med.* **205**, 2703–2710 (2008).
2. Martinez-Moczygemba, M. et al. Pulmonary alveolar proteinosis caused by deletion of the GM-CSFR $\alpha$  gene in the X chromosome pseudoautosomal region 1. *J. Exp. Med.* **205**, 2711–2716 (2008).
3. Suzuki, T. et al. Hereditary pulmonary alveolar proteinosis: pathogenesis, presentation, diagnosis, and therapy. *Am. J. Respir. Crit. Care Med.* **182**, 1292–1304 (2010).
4. Tanaka, T. et al. Adult-onset hereditary pulmonary alveolar proteinosis caused by a single-base deletion in CSF2RB. *J. Med. Genet.* **48**, 205–209 (2011).



5. Suzuki, T. *et al.* Hereditary pulmonary alveolar proteinosis caused by recessive CSF2RB mutations. *Eur. Respir. J.* **37**, 201–204 (2011).
6. Whitsett, J. A., Wert, S. E. & Weaver, T. E. Alveolar surfactant homeostasis and the pathogenesis of pulmonary disease. *Annu. Rev. Med.* **61**, 105–119 (2010).
7. Hawgood, S. & Poulain, F. R. The pulmonary collectins and surfactant metabolism. *Annu. Rev. Physiol.* **63**, 495–519 (2001).
8. Ikegami, M. *et al.* Surfactant metabolism in transgenic mice after granulocyte macrophage-colony stimulating factor ablation. *Am. J. Physiol.* **270**, L650–L658 (1996).
9. Kitamura, T. *et al.* Idiopathic pulmonary alveolar proteinosis as an autoimmune disease with neutralizing antibody against granulocyte/macrophage colony-stimulating factor. *J. Exp. Med.* **190**, 875–880 (1999).
10. Trapnell, B. C., Whitsett, J. A. & Nakata, K. Pulmonary alveolar proteinosis. *N. Engl. J. Med.* **349**, 2527–2539 (2003).
11. Stanley, E. *et al.* Granulocyte/macrophage colony-stimulating factor-deficient mice show no major perturbation of hematopoiesis but develop a characteristic pulmonary pathology. *Proc. Natl Acad. Sci. USA* **91**, 5592–5596 (1994).
12. Dranoff, G. *et al.* Involvement of granulocyte-macrophage colony-stimulating factor in pulmonary homeostasis. *Science* **264**, 713–716 (1994).
13. Robb, L. *et al.* Hematopoietic and lung abnormalities in mice with a null mutation of the common beta subunit of the receptors for granulocyte-macrophage colony-stimulating factor and interleukins 3 and 5. *Proc. Natl Acad. Sci. USA* **92**, 9565–9569 (1995).
14. Nishinakamura, R. *et al.* Mice deficient for the IL-3/GM-CSF/IL-5  $\beta$ c receptor exhibit lung pathology and impaired immune response, while  $\beta$ IL3 receptor-deficient mice are normal. *Immunity* **2**, 211–222 (1995).
15. Bonfield, T. L. *et al.* Peroxisome proliferator-activated receptor- $\gamma$  is deficient in alveolar macrophages from patients with alveolar proteinosis. *Am. J. Respir. Cell Mol. Biol.* **29**, 677–682 (2003).
16. Bonfield, T. L. *et al.* PU.1 regulation of human alveolar macrophage differentiation requires granulocyte-macrophage colony-stimulating factor. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **285**, L1132–L1136 (2003).
17. Seymour, J. F. & Presneill, J. J. Pulmonary alveolar proteinosis: progress in the first 44 years. *Am. J. Respir. Crit. Care Med.* **166**, 215–235 (2002).
18. Shibata, Y. *et al.* GM-CSF regulates alveolar macrophage differentiation and innate immunity in the lung through PU.1. *Immunity* **15**, 557–567 (2001).
19. Thomassen, M. J. *et al.* ABCG1 is deficient in alveolar macrophages of GM-CSF knockout mice and patients with pulmonary alveolar proteinosis. *J. Lipid Res.* **48**, 2762–2768 (2007).
20. Iyonaga, K. *et al.* Elevated bronchoalveolar concentrations of MCP-1 in patients with pulmonary alveolar proteinosis. *Eur. Respir. J.* **14**, 383–389 (1999).
21. Nishinakamura, R. *et al.* The pulmonary alveolar proteinosis in granulocyte macrophage colony-stimulating factor/interleukins 3/5  $\beta$ c receptor-deficient mice is reversed by bone marrow transplantation. *J. Exp. Med.* **183**, 2657–2662 (1996).
22. Kleff, V. *et al.* Gene therapy of  $\beta$ c-deficient pulmonary alveolar proteinosis ( $\beta$ c-PAP): studies in a murine in vivo model. *Mol. Ther.* **16**, 757–764 (2008).
23. Suzuki, T. *et al.* Use of induced pluripotent stem cells to recapitulate pulmonary alveolar proteinosis pathogenesis. *Am. J. Respir. Crit. Care Med.* **189**, 183–193 (2014).
24. Stradling, J. R. & Lane, D. J. Development of secondary polycythaemia in chronic airways obstruction. *Thorax* **36**, 321–325 (1981).
25. Faust, N., Varas, F., Kelly, L. M., Heck, S. & Graf, T. Insertion of enhanced green fluorescent protein into the lysozyme gene creates mice with green fluorescent granulocytes and macrophages. *Blood* **96**, 719–726 (2000).
26. Yoshida, M., Ikegami, M., Reed, J. A., Chroneos, Z. C. & Whitsett, J. A. GM-CSF regulates surfactant Protein-A and lipid catabolism by alveolar macrophages. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **280**, L379–L386 (2001).
27. Lachmann, N. *et al.* Gene correction of human induced pluripotent stem cells repairs the cellular phenotype in pulmonary alveolar proteinosis. *Am. J. Respir. Crit. Care Med.* **189**, 167–182 (2014).
28. Seymour, J. F. *et al.* Mice lacking both granulocyte colony-stimulating factor (CSF) and granulocyte-macrophage CSF have impaired reproductive capacity, perturbed neonatal granulopoiesis, lung disease, amyloidosis, and reduced long-term survival. *Blood* **90**, 3037–3049 (1997).
29. Wu, M. *et al.* Genetically engineered macrophages expressing IFN- $\gamma$  restore alveolar immune function in *scid* mice. *Proc. Natl Acad. Sci. USA* **98**, 14589–14594 (2001).
30. Guillemins, M. *et al.* Alveolar macrophages develop from fetal monocytes that differentiate into long-lived cells in the first week of life via GM-CSF. *J. Exp. Med.* **210**, 1977–1992 (2013).
31. Hashimoto, D. *et al.* Tissue-resident macrophages self-maintain locally throughout adult life with minimal contribution from circulating monocytes. *Immunity* **38**, 792–804 (2013).
32. Yona, S. *et al.* Fate mapping reveals origins and dynamics of monocytes and tissue macrophages under homeostasis. *Immunity* **38**, 79–91 (2013).
33. Schulz, C. *et al.* A lineage of myeloid cells independent of Myb and hematopoietic stem cells. *Science* **336**, 86–90 (2012).
34. van Furth, R. & Cohn, Z. A. The origin and kinetics of mononuclear phagocytes. *J. Exp. Med.* **128**, 415–435 (1968).
35. Godleski, J. J. & Brain, J. D. The origin of alveolar macrophages in mouse radiation chimeras. *J. Exp. Med.* **136**, 630–643 (1972).
36. Matute-Bello, G. *et al.* Optimal timing to repopulation of resident alveolar macrophages with donor cells following total body irradiation and bone marrow transplantation in mice. *J. Immunol. Methods* **292**, 25–34 (2004).
37. Murphy, J., Summer, R., Wilson, A. A., Kotton, D. N. & Fine, A. The prolonged life-span of alveolar macrophages. *Am. J. Respir. Cell Mol. Biol.* **38**, 380–385 (2008).
38. Guth, A. M. *et al.* Lung environment determines unique phenotype of alveolar macrophages. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **296**, L936–L946 (2009).
39. Gautier, E. L. *et al.* Gene-expression profiles and transcriptional regulatory pathways that underlie the identity and diversity of mouse tissue macrophages. *Nature Immunol.* **13**, 1118–1128 (2012).

**Acknowledgements** This work was supported by grants from the NIH (R01 HL085453, R21 HL106134, R01HL118342, 8UL1TR000077-05, AR-47363, DK78392, DK90971), American Thoracic Society Foundation Unrestricted Research Grant, CCHMC Foundation Trustee Grant, Deutsche Forschungsgemeinschaft (DFG; Cluster of Excellence Rebirth; Exc 62/1), the Else Kröner-Fresenius Stiftung, the Eva-Luise Koehler Research Prize for Rare Diseases 2013, and by the Pulmonary Biology Division, CCHMC. Flow cytometric data were acquired within the Research Flow Cytometry Core in the Division of Rheumatology, CCHMC. We thank our hPAP patients and their family members in the United States and internationally for their collaboration; J. Whitsett (CCHMC) and F. McCormack (UCMC) for critical reading of the manuscript; J. Krischer (University of South Florida) and Y. Maeda (CCHMC) for helpful discussions; S. Wert for help with lung histology; and D. Black, K. Link and C. Fox (CCHMC), and S. Brenning and H. Kempf (Hannover Medical School) for their technical help.

**Author Contributions** T.Su., P.A., N.L., S.A., T.M., P.M. and B.C.T. designed research. T.Su., P.A., T.Sa., N.L., C.C., A.S., S.A., B.C. and B.C.T. performed research. T.Su., P.A., T.Sa., N.L., S.A., C.T., T.M., P.M. and B.C.T. analysed data. T.Su., P.A., N.L., P.M., C.L., R.E.W. and B.C.T. wrote the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Microarray data are available at Gene Expression Omnibus under accession number GSE60528. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to B.T. (Bruce.Trapnell@cchmc.org) or T.Su. (Takujii.Suzuki@cchmc.org).

## METHODS

**Mice.** All mice were bred, housed and studied in the Cincinnati Children's Research Foundation Vivarium using protocols approved by the Institutional Animal Care and Use Committee. *Csf2rb* gene-targeted (*Csf2rb*<sup>-/-</sup>) mice<sup>13</sup>, and mice expressing EGFP knocked into the lysozyme M gene (*Lys-M*<sup>GFP</sup> mice)<sup>25</sup>, were all generated previously and backcrossed onto the C57BL/6 background. C57BL/6 mice (referred to as wild type or WT mice) were purchased from Charles River. B6.SJL-*Ptpr*<sup>ca</sup> *Pepc*<sup>b</sup>/BoyJ (CD45.1<sup>+</sup>) mice were from Jackson Laboratory.

**Lung histology and immunohistochemistry.** Animals were killed by intraperitoneal pentobarbital administration and exsanguination by aortic transection. The trachea was exposed by a vertical midline skin incision, cannulated through a small transverse incision in its ventral surface away from the thoracic inlet, inflated with fixative (PBS, pH 7.4, containing 4% paraformaldehyde) under a hydrostatic head of 25 cm and ligated with suture while retracting the cannula to seal the lung under pressure. The sternum and diaphragm were transected sagittally, retracted laterally, and the lungs and heart separated from the chest wall by blunt dissection to avoid puncturing the mediastinal pleura and removed from the chest. The intact tissue block containing the heart, lungs and ligated trachea was submerged in fixative and kept at 4 °C for 24 h. After fixation, the lung lobes were divided, removed from the tissue block, cut into ~2-mm-thick slices along the long axis, washed in cold PBS, dehydrated, embedded in paraffin, and 5-µm-thick sections were cut and stained with haematoxylin and eosin (H&E), periodic acid-Schiff reagent (PAS), or Masson's trichrome as previously described<sup>40</sup>. Immunostaining for surfactant protein B (SP-B) was done by incubating slides with rabbit anti-SP-B polyclonal antibody (diluted 1:500, Seven Hills Bioreagents, Cincinnati, OH) and Vectastain ABC anti-rabbit immunohistochemical horseradish peroxidase kit (Vector Labs, Inc., Burlingame, CA) and counterstaining with haematoxylin as described<sup>41</sup>. To prepare frozen lung sections, the lungs were inflated fixed *in situ* as described above and then the heart and lungs were removed en bloc and cryoprotected by sequential immersion in PBS containing increasing sucrose concentrations (10%, 15% and 20%; 8–12 h, 4 °C, at each concentration). The lungs were then embedded in Tissue-Tek OCT compound (Sakura Finetek, Torrance, CA), frozen and stored at -80 °C until use. Serial 6 µm sections were prepared for immunostaining or evaluation of GFP<sup>+</sup> cells. Lung sections and sedimented lung cells were examined by light microscopy using a Zeiss Axioplan 2 microscope (Zeiss) equipped with AxioVision software (Zeiss).

**Collection, handling and evaluation of bronchoalveolar lavage fluid and cells.** Epithelial lining fluid and non-adherent cells were collected from lung surface of mice by bronchoalveolar lavage (BAL) as described<sup>42</sup> and processed immediately. Briefly, five 1-ml aliquots were instilled and immediately recovered per mouse and combined resulting in a BAL recovery of  $93.9 \pm 1.2\%$  per mouse (BAL recovery data for 10 mice evaluated randomly). The photographs of fresh BAL specimens and the specimen after allowing sediment to be formed by overnight incubation at 4 °C were taken. The turbidity of BAL was determined as described<sup>41</sup>. Briefly, after gently mixing to ensure a homogeneous suspension of BAL, a 250 µl aliquot was diluted into 750 µl PBS and the optical density was measured at a wavelength of 600 nm and multiplying the result by the dilution factor. The total number of BAL cells recovered from each mouse was determined by counting cells in an aliquot of known volume using a haemocytometer and multiplying the result by the total volume of BAL and dividing by the volume of the aliquot used for counting. BAL cytology was evaluated in aliquots (~50,000 cells) after sedimentation (Cytospin, Shandon, Inc.; 500 r.p.m., 7 min, room temperature) onto glass slides and staining with DiffQuick, PAS, or oil red O (all from Fisher Scientific) as described<sup>41</sup>. The cell differential was determined by microscopic examination of DiffQuick stained cells and the total number of alveolar macrophages per mouse was determined by multiplying the percentage of alveolar macrophages in BAL cells by the total number of BAL cells recovered<sup>43</sup>. BAL fluid and cells were separated by low-speed centrifugation (285g, 10 min, room temperature) and stored at -80 °C until use (BAL fluid) or immediately evaluated (referred to as BAL cells) or used to isolate alveolar macrophages (see below). Primary alveolar macrophages were purified by brief adherence of BAL cells to plastic as described<sup>18</sup>. Viability was evaluated by Trypan blue exclusion and was  $\geq 95\%$ .

**ELISA.** The concentration of surfactant protein D (SP-D) in BAL fluid was measured by enzyme-linked immunosorbent assay (ELISA) as we described<sup>41</sup>. The concentration of several cytokines (GM-CSF, M-CSF, MCP-1, IL-1β, IL-6, TNF-α) in BAL fluid and erythropoietin in serum was measured by ELISA (Mouse Quantikine Kits, R&D Systems) as described<sup>1</sup>.

**Quantitative RT-PCR.** Total RNA was isolated from alveolar macrophages using TRIzol Reagent (Life Technologies, Carlsbad, CA) and then used to purify mRNA using RNeasy (Qiagen, Valencia, CA), both as directed by the manufacturers. Purified mRNA was used to synthesize cDNA using the Invitrogen SuperScript III First-Strand Synthesis System (Life Technologies). Standard quantitative RT-PCR (qRT-PCR) was performed as previously described<sup>1</sup> on an Applied Biosystems 7300 Real-Time PCR System (Life Technologies) to measure transcript abundance using

TaqMan oligonucleotide primer sets (all from Life Technologies) (Extended Data Table 1). Expression of target genes was normalized to the expression of 18s RNA. Data for each gene were shown as the fold change of the mean of results for wild-type mice.

**Bone-marrow-derived macrophages (BMDMs).** Bone marrow cells were obtained from 6–8-week-old WT, *Csf2rb*<sup>-/-</sup>, or *Lys-M*<sup>GFP</sup> mice by isolating and flushing tibias and femurs with DMEM (Life Technologies) containing 10% heat-inactivated FBS, 50 U ml<sup>-1</sup> penicillin, and 50 µg ml<sup>-1</sup> streptomycin. After red blood cells were removed with BD Pharm Lyse (BD Biosciences), mononuclear cells were isolated by centrifugation on Ficoll-Paque (GE Healthcare) at room temperature for 30 min, washed, re-suspended in DMEM containing 10% heat-inactivated FBS, 50 U ml<sup>-1</sup> penicillin, 50 µg ml<sup>-1</sup> streptomycin, 10 ng ml<sup>-1</sup> GM-CSF and 5 ng ml<sup>-1</sup> M-CSF (both from R&D Systems), seeded into plastic dishes (Falcon) at a density of  $\sim 27 \times 10^6$  cells per 10 cm dish (1 per mouse) and cultured overnight at 37 °C in a humidified environment containing 5% CO<sub>2</sub>. The next day, non- or weakly-adherent cells were recovered, transferred to a new dish and cultured under the conditions just described to permit differentiation and expansion of macrophages; firmly adherent cells were discarded. After 2 days the culture medium was changed and after 5 days from seeding, adherent bone-marrow-derived macrophages were gently washed with PBS, harvested by brief exposure to trypsin-EDTA (Life Technologies), washed, and used for experiments. The cell purity was high as indicated by the percentage of CD68<sup>+</sup> and F4/80<sup>+</sup> cells ( $96.6 \pm 0.3\%$ ,  $95.4 \pm 1.3\%$ , respectively, not shown).

Some experiments used lineage-negative (Lin<sup>-</sup>) c-Kit<sup>+</sup> Sca-1<sup>+</sup> (LSK) cells which were obtained from mouse bone marrow as described<sup>44</sup>. Briefly, bone marrow from 6–8-week-old WT or *Csf2rb*<sup>-/-</sup> mice was collected as above and lineage depleted with biotinylated lineage antibodies CD5 (53-7.3), CD8a (53-6.7), CD45R/B220 (RA3-6B2), CD11b (M1/70), Gr-1 (RB6-8C5), and TER-119 (TER-119) (BD Bioscience), and magnetic beads (Dynabeads sheep anti-rat IgG) (Life Technologies). After removing lineage-positive cells, the remaining cells were stained with 7-ADD, FITC-Streptavidin (BD Biosciences) and antibodies to Sca-1 (D7) and c-Kit (2B8) (BD Biosciences). Then, Lin<sup>-</sup> c-Kit<sup>+</sup> Sca-1<sup>+</sup> 7-ADD<sup>-</sup> cells were isolated by cell sorting on a FACSaria (BD Biosciences) and used immediately in experiments. Cell morphology was confirmed by DiffQuick Staining of sedimented cells (Cytospin, Shandon) and viability was measured by Trypan blue exclusion as described<sup>18</sup> and found to be  $\geq 95\%$ . In some experiments, cells were immunostained for CD68 (FA-11) (AbD Serotec), counterstained with DAPI as described<sup>41</sup>, and examined by light microscopy using a Zeiss Axioplan 2 microscope (Zeiss) equipped with AxioVision software (Zeiss).

**Colony forming cell (CFC) assay.** BMDMs or Lin<sup>-</sup> bone marrow cells were evaluated for the presence of haematopoietic progenitors capable of forming colonies in semisolid medium in response to cytokine stimulation as previously described<sup>45</sup>. Briefly, fresh Lin<sup>-</sup> bone marrow cells or BMDMs after induced differentiation into macrophages for 5 days were seeded into standard mouse methylcellulose media supplemented with insulin, transferrin, SCF, IL-3, IL-6 and erythropoietin (HSC007, R&D Systems, Minneapolis, MN). After 7 days in culture, colonies of  $\geq 50$  cells were visible and were examined morphologically using whole-plate stack images acquired using an AXIO-Z1 microscope and AXIO-vision software (Zeiss, Jena, Germany) to identify and enumerate burst-forming erythroid progenitors (BFU-E), colony-forming myeloid progenitors (CFU-GM) and the multi-potential progenitors (CFU-GEMM).

**Surfactant clearance assay.** BMDMs were evaluated functionally to demonstrate their ability to clear human surfactant as we previously reported<sup>41</sup>. Briefly, BMDMs from either WT or *Csf2rb*<sup>-/-</sup> mice were seeded into 12-well plates ( $4 \times 10^5$  cells per well) in DMEM, 10% FBS, 10 ng ml<sup>-1</sup> GM-CSF, 5 ng ml<sup>-1</sup> M-CSF. Human surfactant recovered by lavage of a patient with PAP was added to the media and cells were incubated for 24 h to permit surfactant uptake into cells and then washed to remove extracellular surfactant. Cells were incubated for 24 h to permit macrophages to clear internalized surfactant. Cells collected before, immediately after surfactant exposure or 24 h after the completion of surfactant exposure were sedimented onto slides by cytocentrifugation (Shandon), stained with oil red O, and counterstained with haematoxylin. Oil red O staining was evaluated in  $\geq 10$  random  $20\times$  microscopic fields for each sample as described<sup>41</sup>.

**Pulmonary macrophage transplantation (PMT).** BMDMs or LSK cell-derived macrophages were administered directly into the lungs of 8-week-old mice using a relatively non-invasive endotracheal instillation method described previously<sup>46</sup>. Briefly, mice received light anaesthesia by isoflurane inhalation and were suspended on a flat board by a rubber band across the upper incisors and placed in a semi-recumbent (45°) position with the ventral surface and rostrum facing upwards. Using a curved blade Kelly forceps, the tongue was gently and partially retracted rostrally, and 50 µl of PBS containing the macrophages to be administered was placed in the back of the oral cavity using a micropipette. The PBS and cells were inhaled into the lungs by subsequent respiratory efforts under direct visualization. Mice were then observed while recovering from anaesthesia to ensure continued



retention of the administered fluid and cells and then returned to their cages for routine care and handling. Because the efficacy of PMT at a dose of two million macrophages was optimal, this dose given as one administration was used throughout the study except where noted (Extended Data Tables 2 and 3). Age-matched mice were used in all experiments to control for the degree of lung disease severity.

**Flow cytometry.** BAL cells were purified by centrifugation on Percoll to remove surfactant and debris<sup>31</sup>. BAL cells or BMDMs were immunostained to detect CD115 (AFS98), F4/80 (BM8) (eBioscience), CD3 (145-2C11), CD11b (M1/70), CD11c (HL3), CD16/32 (2.4G2), CD19 (1D3), CD64 (X54-5/7.1.1), CD131 (JORO50), Ly6G (1A8), CD45.1 (A20), CD45.2 (104), MHC class II (I-A/I-E) (M5/114.15.2), SiglecF (E50-2440) (BD Biosciences), CD14 (Sa14-2) (BioLegend), CD68 (FA-11) (AbD Serotec), and MerTK (108928) (R&D Systems), as previously described<sup>1</sup>, evaluated by flow cytometry using a FACSCaliber or FACSCanto flow cytometers (both from BD Biosciences, San Jose, CA), and the results were analysed using CellQuest, FACSDiva (BD Biosciences), or FlowJo software (Tree Star). For intracellular staining of CD68, Leucoperm (AbD Serotec) was used as directed by the manufacturer.

**Quantification of CD131<sup>+</sup> alveolar macrophages.** The percentage of BAL cells expressing the GM-CSF receptor  $\beta$ -subunit was determined by immunostaining. Briefly, aliquots of BAL cells were sedimented onto glass slides and incubated (10 min, room temperature) in fixative (PBS containing 4% paraformaldehyde), washed with PBS and incubated (4 °C, overnight) with anti-mouse GM-CSF receptor- $\beta$  (CD131) antibody (sc-678) (Santa Cruz) diluted 1:400 in PBST (PBS containing 2.5% (w/v) Triton X-100 and 5% (v/v) goat serum). After incubation, slides were rinsed five times in PBST and incubated (room temperature, 1 h) with the secondary detection antibody (Alexa-Fluor-594-conjugated, anti-rabbit IgG (Life Technologies)) and counterstained with 4', 6-diamidino-2-phenylindole dihydrochloride (DAPI) (Vector Labs, Burlington, CA). Cells were examined using a Zeiss Axioplan 2 microscope (Zeiss) equipped with AxioVision software (Zeiss). The percentage of CD131<sup>+</sup> BAL cells was determined by first counting the CD131<sup>+</sup> and DAPI<sup>+</sup> cells in five (or more) random 20 $\times$  microscopic fields for each BAL sample. Then, the number of CD131<sup>+</sup> cells in each field was divided by the number of DAPI<sup>+</sup> cells in the same field and results for all fields examined were averaged and multiplied by 100. The total number of CD131<sup>+</sup> cells per mouse was calculated by multiplying the percentage of CD131<sup>+</sup> cells by the total number of BAL cells recovered from each mouse.

**STAT5 phosphorylation index assay.** GM-CSF bioactivity in BAL fluid and GM-CSF receptor function in transduced or WT macrophages was evaluated by measuring GM-CSF-stimulated phosphorylation of STAT5 in BMDMs or LSK cell-derived macrophages using anti-phospho STAT5 antibody (47/Stat5(pY694)) (BD Biosciences) by flow cytometry as previously reported<sup>1</sup>. The STAT5 phosphorylation index (STAT5-PI) was calculated as the mean fluorescence intensity of phosphorylated STAT5 staining in GM-CSF-stimulated cells minus that of non-stimulated cells, divided by that of non-stimulated cells, and multiplied by 100. In experiments to quantify GM-CSF bioactivity, WT BMDMs were incubated in BAL fluid containing anti-GM-CSF (22E9, eBioscience) or isotype control antibody (10  $\mu$ g ml<sup>-1</sup>) for 30 min and then evaluated.

**Evaluation of macrophage proliferation.** *In vitro mixed-cell proliferation assay.* CD45.1<sup>+</sup> WT LSK-derived cells and CD45.2<sup>+</sup> *Csf2rb*<sup>-/-</sup> LSK-derived cells were isolated, seeded into dishes at an initial ratio of 1:3, respectively, and cultured in DMEM containing 10% bovine calf serum, 1% penicillin/streptomycin, GM-CSF (10 ng ml<sup>-1</sup>) and M-CSF (5 ng ml<sup>-1</sup>). Cells were collected at 1, 7, 14 and 18 days, immunostained with anti-murine CD45.1, anti-CD45.2 and evaluated by flow cytometry to determine the percentage of each cell type at these times.

*In vivo evaluation of transplanted macrophage proliferation.* Frozen lung tissue sections were immunostained with anti-Ki67 antibody (Roche) and examined using a Zeiss Axioplan 2 microscope (Zeiss). The percentage of proliferating PMT-derived cells was determined by enumerating GFP<sup>+</sup> Ki67<sup>+</sup> cells among total GFP<sup>+</sup> cells in  $\geq 7$  random 20 $\times$  microscopic fields for each sample. To confirm the specificity of Ki67 immunostaining, paraffin-embedded sections or WT alveolar macrophages isolated by BAL and adherence were also stained with Ki67 and examined by light microscopy.

**Western blotting.** Detection of GM-CSF receptor- $\beta$  and actin by western blotting was done as previously described<sup>1</sup> with the following modifications. Briefly, primary alveolar macrophages (0.5  $\times 10^6$  per condition) or cultured BMDMs (1  $\times 10^6$  per condition) were collected by low-speed centrifugation (285g, 4 °C, 10 min) and the pellets incubated on ice for 30 min in 200  $\mu$ l of lysis buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 1% (v/v) nonidet p-40, 0.5% (w/v) sodium deoxycholate, 0.1% (w/v) sodium dodecyl sulphate (SDS), 0.004% (w/v) sodium azide) containing 2% (v/v) proteinase inhibitor cocktail (phenyl-methyl-sulphonyl-fluoride and sodium orthovanadate; Santa Cruz). Insoluble debris was removed by centrifugation at 10,000g, 4 °C, 15 min and the supernatant transferred to a clean polypropylene tube. An equal volume of Laemmli sample loading buffer (Bio-Rad, CA) was added and the tubes were capped tightly, vortexed briefly, boiled for 5 min, and separated by electrophoresis on SDS-polyacrylamide gradient (4–12%) gels

(Invitrogen) under reducing conditions. Separated proteins were transferred to PVDF membranes by electro-blotting, incubated in blotting solution (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 5% (w/v) non-fat dry milk (Kroger, Cincinnati, OH), 0.1% (v/v) Tween 20; 4 °C, overnight) to block non-specific binding. Diluted primary detection antibody (see below) was added and the membranes were incubated for 2 h at room temperature and then washed in TBST (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.1% (v/v) Tween 20). Membranes were then incubated with the secondary HRP-conjugated detection antibody in blotting solution for 1 h at room temperature and then washed as above and then incubated with ECL-Plus (GE Healthcare) as directed by the manufacturer. Anti-mouse GM-CSF receptor- $\beta$  antibody (sc-678) (Santa Cruz) diluted 1:500 and anti-actin (sc-1616) (Santa Cruz) diluted 1:1,000 were used for primary antibodies.

**Haematological analysis.** Blood was obtained from the superior vena cava from mice and 20  $\mu$ l was used to measure complete blood counts on a fully automated Hemavet 850 (Drew Scientific). Data for the precision and linearity of measurements made with the Hemavet850 can be found online at [http://www.drewscientific.com/product\\_hemavet850.htm](http://www.drewscientific.com/product_hemavet850.htm).

**Microarray analysis.** Alveolar macrophages were obtained from age-matched mice (three per condition) and analysed individually as follows. Total RNA was isolated as described above and microarray analysis was performed using the Mouse Gene 1.0 ST Array (Affymetrix, Santa Clara, CA) in the CCHMC Affymetrix Core using standard procedures as described<sup>47</sup>. Data (available at Gene Expression Omnibus accession GSE60528) were analysed using the Affymetrix package in the R statistical programming language (Bioconductor; <http://www.bioconductor.org>). Probes were corrected for background using the Microarray Analysis Suite algorithm, quantile normalized, and probe sets were summarized using the average difference of perfect matches only. Differential expression tests were performed using significance analysis of microarrays<sup>48</sup> with Benjamini–Hochberg correction for multiple testing<sup>49</sup>. Significant gene lists were selected with a  $\Delta$  that constrained the false discovery rate to less than 10%. Cluster dendrogram was generated from unsupervised hierarchical clustering analysis of microarray data from probes for all 28,853 genes represented on the chip (Spearman correlation; 3 mice per group). In Venn diagrams, numbers of genes for which expression was altered in alveolar macrophages from *Csf2rb*<sup>-/-</sup> compared to WT mice (WT $\rightarrow$ KO) or PMT-treated compared to untreated *Csf2rb*<sup>-/-</sup> mice (KO $\rightarrow$ KO+PMT) were shown. Only genes with statistically significant changes (false detection rate <10%) of at least twofold were marked as increased (up arrows) or decreased (down arrows). The numbers of genes for which expression was disrupted in *Csf2rb*<sup>-/-</sup> mice and normalized by PMT (or unchanged in both comparisons) is shown in the overlap regions. In gene ontology analysis, data show the coordinate increases (red) or decreases (blue) in expression of genes in all gene sets significant at or below a false detection rate of 10% calculated by the Gene Set Test with correction for multiple testing.

**Lentiviral vectors, LSK-cell transduction, and differentiation and expansion of transduced macrophages.** Gene transfer vectors were constructed using routine methods<sup>44</sup> from the vector backbone of (Ery-GFP), a human immunodeficiency virus-based, self-inactivating (SIN) lentiviral vector (LV) harbouring a 398-bp U3 deletion eliminating the strong viral promoter/enhancer element<sup>50</sup>. GM-R-LV contains a chimaeric transgene comprised of the human elongation factor 1- $\alpha$  (ELF1- $\alpha$ ) promoter (a 1,189-bp fragment containing intron 1 ending 20 bp upstream of the ATG codon isolated from the pEF-BOS plasmid<sup>51</sup>) followed by the mouse *Csf2rb* cDNA (nucleotides -80 to 2,691, GenBank accession number M34397.1) located 3' of the lentiviral central poly-purine tract and followed by an internal ribosome binding site (IRES) and then an enhanced green fluorescent protein (GFP) transgene (Fig. 5a). GFP-LV is a lentiviral vector of similar design except that the *Csf2rb* and IRES were omitted and the GFP transgene is driven from the ELF1 $\alpha$  promoter (Fig. 5a). Both vectors contain a viral splice donor site, packaging sequence, splice acceptor site, and central polypurine tract (cPPT) 5' of the ELF1 $\alpha$  promoter and a woodchuck hepatitis post-transcriptional regulatory element (WPRE) (nucleotides 1093 to 1684; GenBank accession number J04514)<sup>52</sup> located 3' of the GFP stop codon as described<sup>50</sup>. Lentiviral vectors were produced by transient transfection as vesicular stomatitis virus-G (VSV-G) virions, concentrated, and titred as described<sup>44</sup>. *Csf2rb*<sup>-/-</sup> LSK cells were isolated, transduced and expanded as described<sup>53</sup> except that transductions were done at a multiplicity of infection (MOI) of 20 for two 12-h periods, IL-11 was omitted, GM-CSF (10 ng ml<sup>-1</sup>) and M-CSF (5 ng ml<sup>-1</sup>) were included, SCF and Flt-3 ligand were sequentially reduced (50, 1, 0 ng ml<sup>-1</sup>), and IL-3 was present early.

Transduction, expansion and differentiation of LSK cells into gene-corrected macrophages was done by adjusting the cytokine 'cocktail' mixture to optimize the culture conditions for each of four sequential stages, which included: (1) LSK transduction: murine SCF (R&D) 50 ng ml<sup>-1</sup>, mIL-3 (PeproTech) 10 ng ml<sup>-1</sup>, hFlt3-L (PeproTech) 50 ng ml<sup>-1</sup> and GM-CSF (R&D) 10 ng ml<sup>-1</sup>, culture time of two 12-h periods; (2) progenitor expansion: mSCF 50 ng ml<sup>-1</sup>, hFlt3-L 50 ng ml<sup>-1</sup> and GM-CSF 10 ng ml<sup>-1</sup>, culture time of 4 days; (3) macrophage lineage commitment: mSCF 1 ng ml<sup>-1</sup>, hFlt3-L 1 ng ml<sup>-1</sup>, GM-CSF 10 ng ml<sup>-1</sup>, M-CSF (R&D) 5 ng ml<sup>-1</sup>, culture

time of 3 days; and (4) macrophage differentiation: GM-CSF  $10 \text{ ng ml}^{-1}$  and M-CSF  $5 \text{ ng ml}^{-1}$ , culture time of 4 days. StemSpan (STEMCELL Technologies) containing 2% FBS, 1% penicillin/streptomycin,  $10 \text{ mM dNTP}$ , and low-density lipoprotein was used as the culture medium for the LSK transduction and DMEM with 10% FBS,  $50 \text{ U ml}^{-1}$  penicillin and  $50 \mu\text{g ml}^{-1}$  streptomycin was used for all other stages. Phenotype markers (F4/80, CD11b, CD11c) were analysed by flow cytometry at each stage to monitor macrophage differentiation. Only adherent macrophages at the end of this procedure were used for PMT.

**Localization of PMT-derived cells after transplantation.** Several approaches were used to identify and localize PMT-derived cells within the lung parenchyma and in different organs.

**Intra-pulmonary localization of PMT-derived cells.** CD131 immunostaining and fluorescence microscopy or flow cytometry was used to detect and quantify transplantation-derived donor macrophages among BAL cells from the lungs of *Csf2rb*<sup>-/-</sup> mice that previously received PMT of WT (C57BL/6) BMDMs, Lys-M<sup>GFP</sup> BMDMs, CD45.1<sup>+</sup> WT BMDMs, or GM-R-GFP-LV *Csf2rb* gene-corrected *Csf2rb*<sup>-/-</sup> LSK-derived macrophages.

To localize PMT-derived macrophages to intra-alveolar space or interstitium of the lung, frozen lung sections from mice that received PMT of Lys-M<sup>GFP</sup> BMDMs 1 month or 1 year earlier were immunostained with CD68, counterstained with DAPI (Vector Labs) and evaluated by fluorescence microscopy to identify macrophages, PMT-derived cells, and nucleated cells, respectively. PMT-derived macrophages (that is, GFP<sup>+</sup> CD68<sup>+</sup> cells) located within the intra-alveolar space or the interstitium were then enumerated. To eliminate the possibility of any interference from non-specific auto-fluorescence of alveolar macrophages, paraffin-embedded lung sections from these mice were immunostained with anti-GFP antibody (Life Technologies) and examined by light microscopy to enumerate immunohistochemically marked macrophages located within the intra-alveolar space or interstitium.

**Organ-specific localization of PMT-derived cells.** In one approach, *Csf2rb*<sup>-/-</sup> mice received PMT of Lys-M<sup>GFP</sup> BMDMs and 1 year later, cells isolated from the BAL, blood, bone marrow, and spleen were evaluated by flow cytometry to detect GFP<sup>+</sup> cells as a marker for PMT-derived cells.

In a second approach, CD45.2<sup>+</sup> *Csf2rb*<sup>-/-</sup> mice received PMT of CD45.1<sup>+</sup> BMDMs and 1 year later, cells isolated from the BAL, blood, bone marrow and spleen were evaluated by flow cytometry to detect CD45.1<sup>+</sup> cells as a marker for PMT-derived cells.

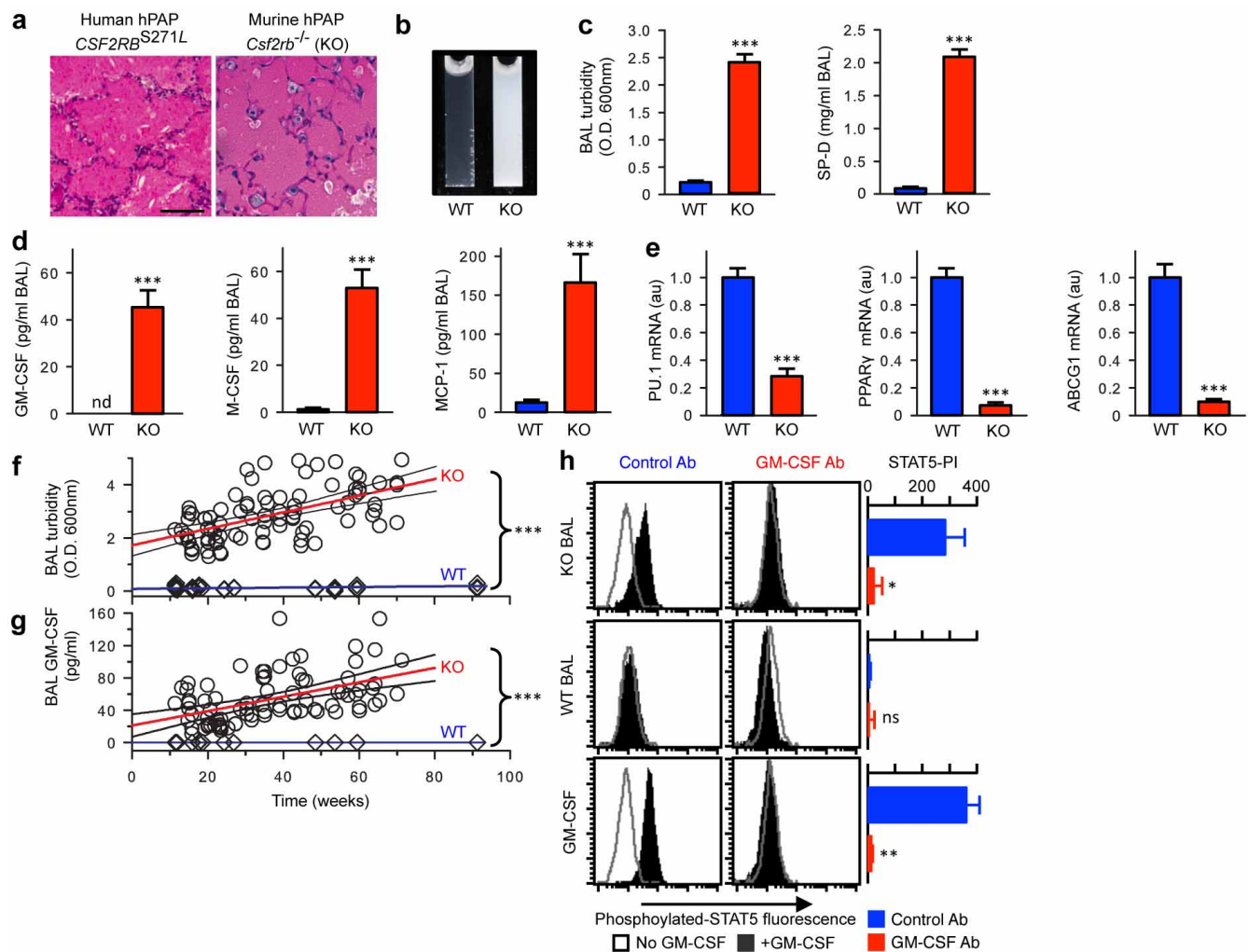
In a third approach, *Csf2rb*<sup>-/-</sup> mice received PMT of Lys-M<sup>GFP</sup> BMDMs and 1 year later, DNA was extracted from the BAL cells (lung), blood leukocytes, bone marrow cells, and spleen using a DNeasy Blood & Tissue Kit (Qiagen). Organ-specific DNA was subjected to PCR amplification using oligonucleotide primers (Extended Data Table 1) specific for the Lys-M<sup>GFP</sup> knock-in transgene or the unmodified Lysozyme M gene to detect PMT-derived and endogenous cells, respectively, as previously reported<sup>25</sup>.

A fourth approach was conducted using a specific operating procedure (TSL 6-13) and Good Laboratory Practice (GLP) conditions within the CCHMC Translational Core Laboratory. Here, DNA was extracted from the BAL cells (lung), blood leukocytes, bone marrow cells, and spleen of *Csf2rb*<sup>-/-</sup> mice that had received PMT of GM-R-GFP-LV *Csf2rb* gene-corrected, *Csf2rb*<sup>-/-</sup> LSK-derived macrophages 1 year earlier and subjected to quantitative PCR amplification with oligonucleotide primers specific for the R-U5 of GM-R-GFP-LV using Applied Biosystems ABI7900HT Fast Real-Time PCR System (Life Technologies). The number of GM-R-GFP-LV vector copies per microgram of organ-specific DNA was quantified and normalized to the level of mouse apolipoprotein B gene as described previously<sup>54</sup>.

**Statistical analysis.** Numeric data were evaluated for normality and variance using the Shapiro-Wilk and Levene median tests, respectively, and presented as mean  $\pm$  s.e.m. (parametric data) or median and interquartile range (nonparametric data). Statistical comparisons were made with Student's *t*-test, one-way analysis of variance, or Kruskal-Wallis rank-sum test as appropriate; post-hoc pairwise multiple comparison procedures were done using the Student-Newman-Keuls or Dunn's method as appropriate. *P* values of  $\leq 0.05$  were considered to indicate statistical significance. Based on the use of BAL turbidity (the primary outcome variable for efficacy) measured 2 months after PMT of WT BMDMs into *Csf2rb*<sup>-/-</sup> mice and compared to age-matched, untreated *Csf2rb*<sup>-/-</sup> mice, 6 mice per group had a power 0.8 to detect a difference of 1.4 OD 600 nm using a two-tailed Student's *t*-test and a *P* value of 0.05. All studies used male and female mice by randomly assigning mice housed in the same cage to separate experimental groups but without formal randomization or blinding. Results from all mice were included in the final analysis without exclusion. Analyses, including Kaplan-Meier survival analysis, were performed with SigmaPlot, Version 12.5 (Systat Software, San Jose, CA). All experiments were repeated at least twice, with similar results.

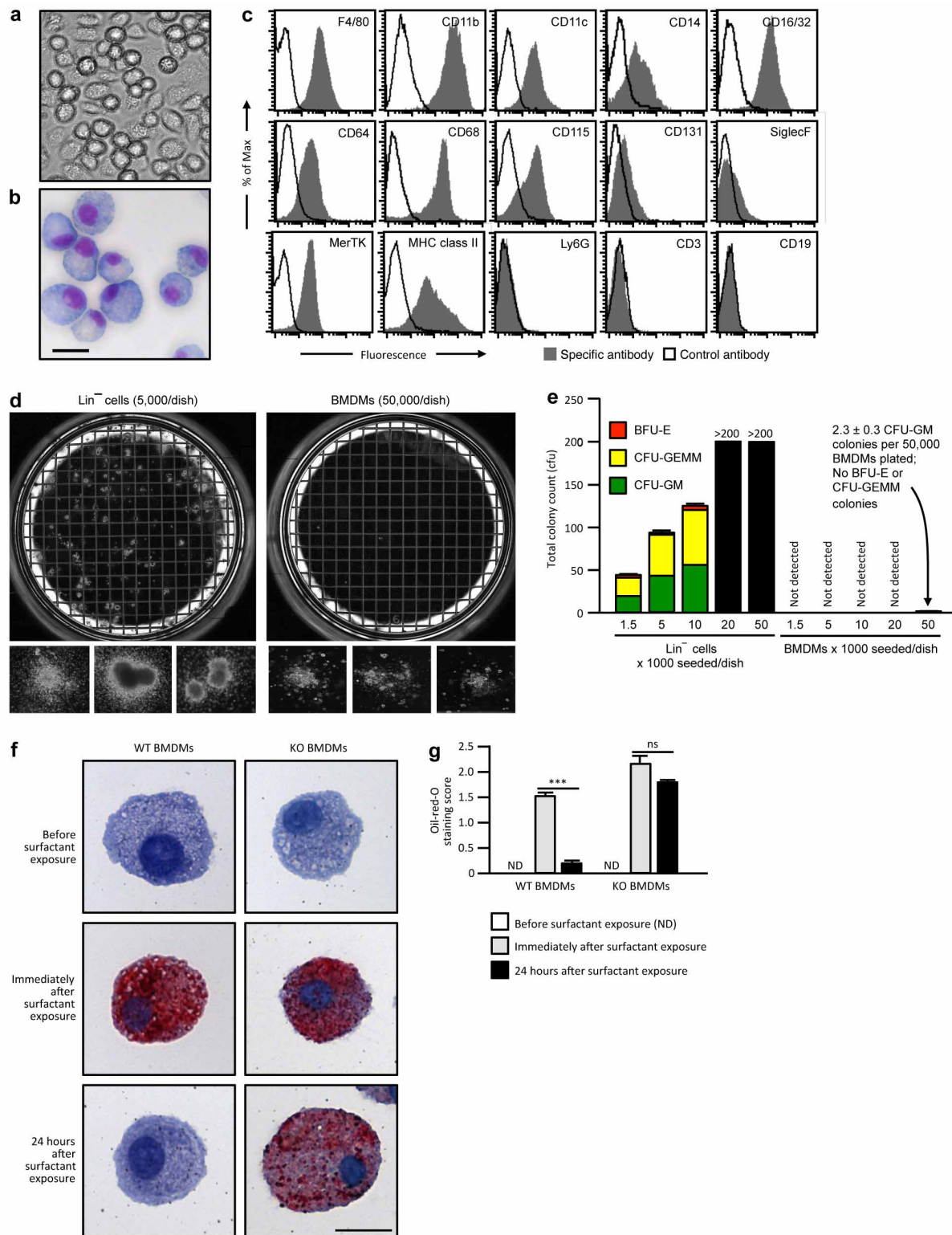
40. Wert, S. E. *et al.* Increased metalloproteinase activity, oxidant production, and emphysema in surfactant protein D gene-inactivated mice. *Proc. Natl Acad. Sci. USA* **97**, 5972–5977 (2000).
41. Sakagami, T. *et al.* Patient-derived granulocyte/macrophage colony-stimulating factor autoantibodies reproduce pulmonary alveolar proteinosis in nonhuman primates. *Am. J. Respir. Crit. Care Med.* **182**, 49–61 (2010).
42. LeVine, A. M., Reed, J. A., Kurak, K. E., Cianciolo, E. & Whitsett, J. A. GM-CSF-deficient mice are susceptible to pulmonary group B streptococcal infection. *J. Clin. Invest.* **103**, 563–569 (1999).
43. Berclaz, P. Y. *et al.* GM-CSF regulates a PU.1-dependent transcriptional program determining the pulmonary response to LPS. *Am. J. Respir. Cell Mol. Biol.* **36**, 114–121 (2007).
44. Perumbeti, A. *et al.* A novel human gamma-globin gene vector for genetic correction of sickle cell anemia in a humanized sickle mouse model: critical determinants for successful correction. *Blood* **114**, 1174–1185 (2009).
45. Kaufman, D. S., Hanson, E. T., Lewis, R. L., Auerbach, R. & Thomson, J. A. Hematopoietic colony-forming cells derived from human embryonic stem cells. *Proc. Natl Acad. Sci. USA* **98**, 10716–10721 (2001).
46. Walters, D. M., Breyse, P. N. & Wills-Karp, M. Ambient urban Baltimore particulate-induced airway hyperresponsiveness and inflammation in mice. *Am. J. Respir. Crit. Care Med.* **164**, 1438–1443 (2001).
47. Maeda, Y. *et al.* Kras(G12D) and Nkx2-1 haploinsufficiency induce mucinous adenocarcinoma of the lung. *J. Clin. Invest.* **122**, 4388–4400 (2012).
48. Tusher, V. G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl Acad. Sci. USA* **98**, 5116–5121 (2001).
49. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).
50. Richard, E. *et al.* Gene therapy of a mouse model of protoporphyria with a self-inactivating erythroid-specific lentiviral vector without preselection. *Mol. Ther.* **4**, 331–338 (2001).
51. Mizushima, S. & Nagata, S. pEF-BOS, a powerful mammalian expression vector. *Nucleic Acids Res.* **18**, 5322 (1990).
52. Zufferey, R., Donello, J. E., Trono, D. & Hope, T. J. Woodchuck hepatitis virus posttranscriptional regulatory element enhances expression of transgenes delivered by retroviral vectors. *J. Virol.* **73**, 2886–2892 (1999).
53. Arumugam, P. I. *et al.* The 3' region of the chicken hypersensitive site-4 insulator has properties similar to its core and is required for full insulator activity. *PLoS ONE* **4**, e6995 (2009).
54. Arumugam, P. I. *et al.* Improved human  $\beta$ -globin expression from self-inactivating lentiviral vectors carrying the chicken hypersensitive site-4 (cHS4) insulator element. *Mol. Ther.* **15**, 1863–1871 (2007).





**Extended Data Figure 1 | Validation of *Csf2rb*<sup>-/-</sup> mice as an authentic model of human hPAP.** **a**, Typical lung pathology showing surfactant-filled alveoli with well-preserved septa in a child homozygous for *CSF2RB*<sup>S271L</sup> mutations and identical pulmonary histopathology in a *Csf2rb*<sup>-/-</sup> mouse. PAS stain. Scale bar, 100  $\mu$ m. **b**, Photographs of 'milky'-appearing BAL from a 14-month-old *Csf2rb*<sup>-/-</sup> mouse and normal-appearing BAL from an age-matched WT mouse (representative of  $n = 6$  mice per group). **c**, Increased BAL turbidity and SP-D concentration in 4-month-old *Csf2rb*<sup>-/-</sup> mice compared to age-matched WT mice. **d**, BAL fluid biomarkers of hPAP (GM-CSF, M-CSF and MCP-1) are increased in 4-month-old *Csf2rb*<sup>-/-</sup> mice compared to age-matched WT mice. **e**, Alveolar macrophage biomarkers (*PU.1*, *Pparg*, *Abcg1* mRNA) are reduced in 4-month-old *Csf2rb*<sup>-/-</sup> compared to age-matched WT mice. **f**, Progressive increase in BAL turbidity in *Csf2rb*<sup>-/-</sup>

mice but not age-matched WT mice (linear regression: *Csf2rb*<sup>-/-</sup>, slope =  $0.1271 \pm 0.16$  ( $r^2$ , 0.311); WT, slope =  $0.031 \pm 0.005$ ). **g**, Progressive increase in BAL fluid GM-CSF level in *Csf2rb*<sup>-/-</sup> mice but not age-matched WT mice (linear regression: *Csf2rb*<sup>-/-</sup>, slope =  $0.89 \pm 0.016$  ( $r^2$ , 0.249); WT, slope = 0). **h**, GM-CSF bioactivity in BAL fluid from 10-month-old *Csf2rb*<sup>-/-</sup> or WT mice (or 1 ng ml<sup>-1</sup> murine GM-CSF) measured in the presence of anti-GM-CSF antibody (GM-CSF Ab) or isotype control (Control Ab) using the GM-CSF-stimulated STAT5 phosphorylation index (STAT5-PI) assay. Data are mean  $\pm$  s.e.m. of  $n = 7$  mice per group (c-e),  $n = 4$  (h) or symbols representing individual WT ( $n = 38$ ) or *Csf2rb*<sup>-/-</sup> ( $n = 84$ ) mice and regression fit  $\pm$  95% CI (f-g). \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ . ns, not significant.

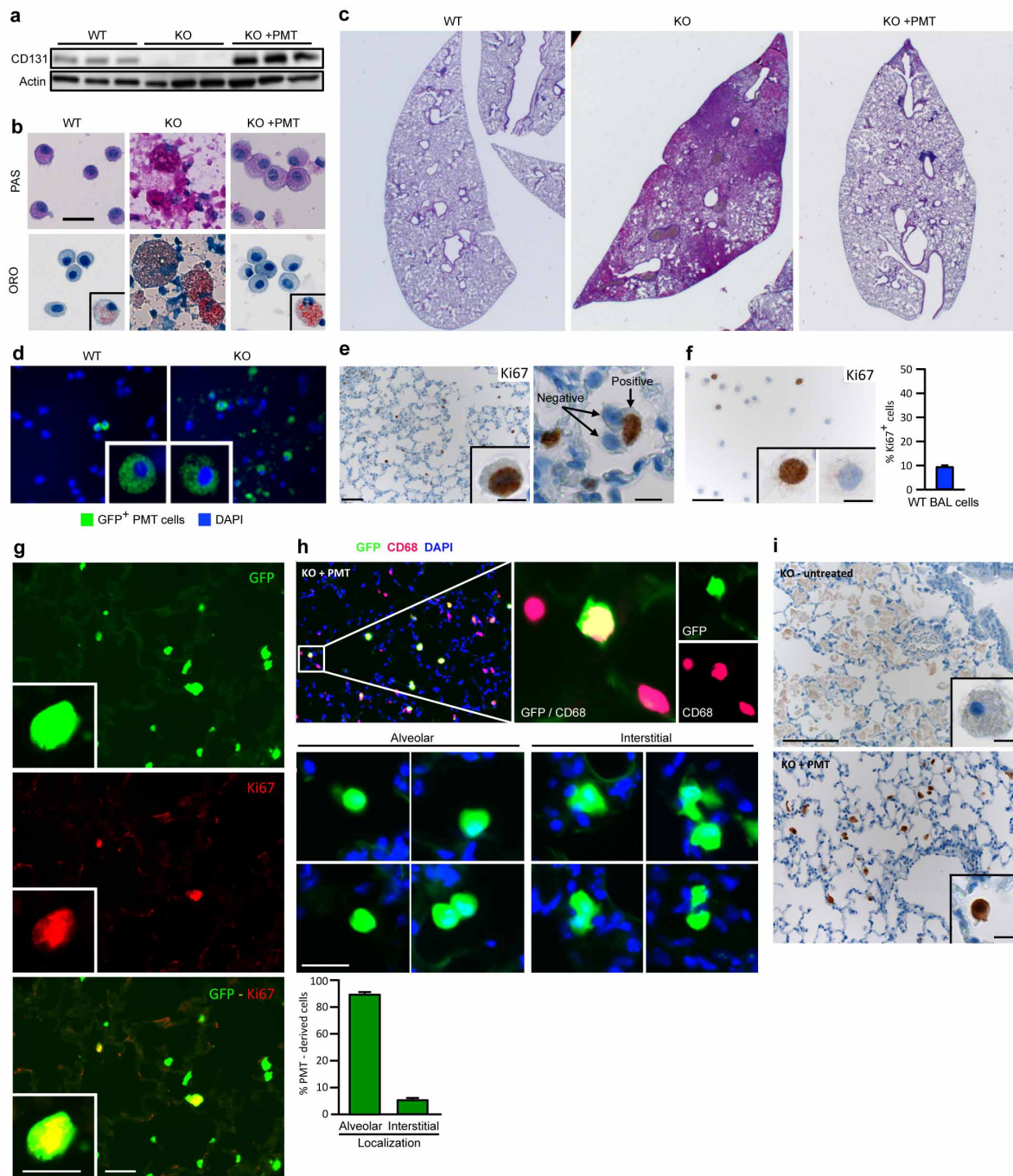


### Extended Data Figure 2 | Characterization of BMDMs before PMT.

**a**, **b**, Photomicrographs of WT BMDMs before transplantation phase-contrast (**a**) or DiffQuick staining (**b**) (representative of  $n = 7$  BMDM preparations). Scale bar, 20  $\mu\text{m}$ . **c**, Flow cytometry evaluation of cell-surface phenotypic markers on WT BMDMs before PMT. **d**, Photographs of methylcellulose cultures of  $\text{Lin}^-$  cells (5,000 per dish) from bone marrow (left) and BMDMs (50,000 per dish) prepared as described in the Methods (right) and typical colonies (below) (representative  $n = 3$  per condition). **e**, Colony counts of BFU-E, CFU-GEMM and CFU-GM showing that BMDMs contained  $<0.005\%$  CFU-GM and no BFU-E or CFU-GEMM progenitors, corresponding to 93 CFU-GM per dose of BMDMs administered ( $n = 3$

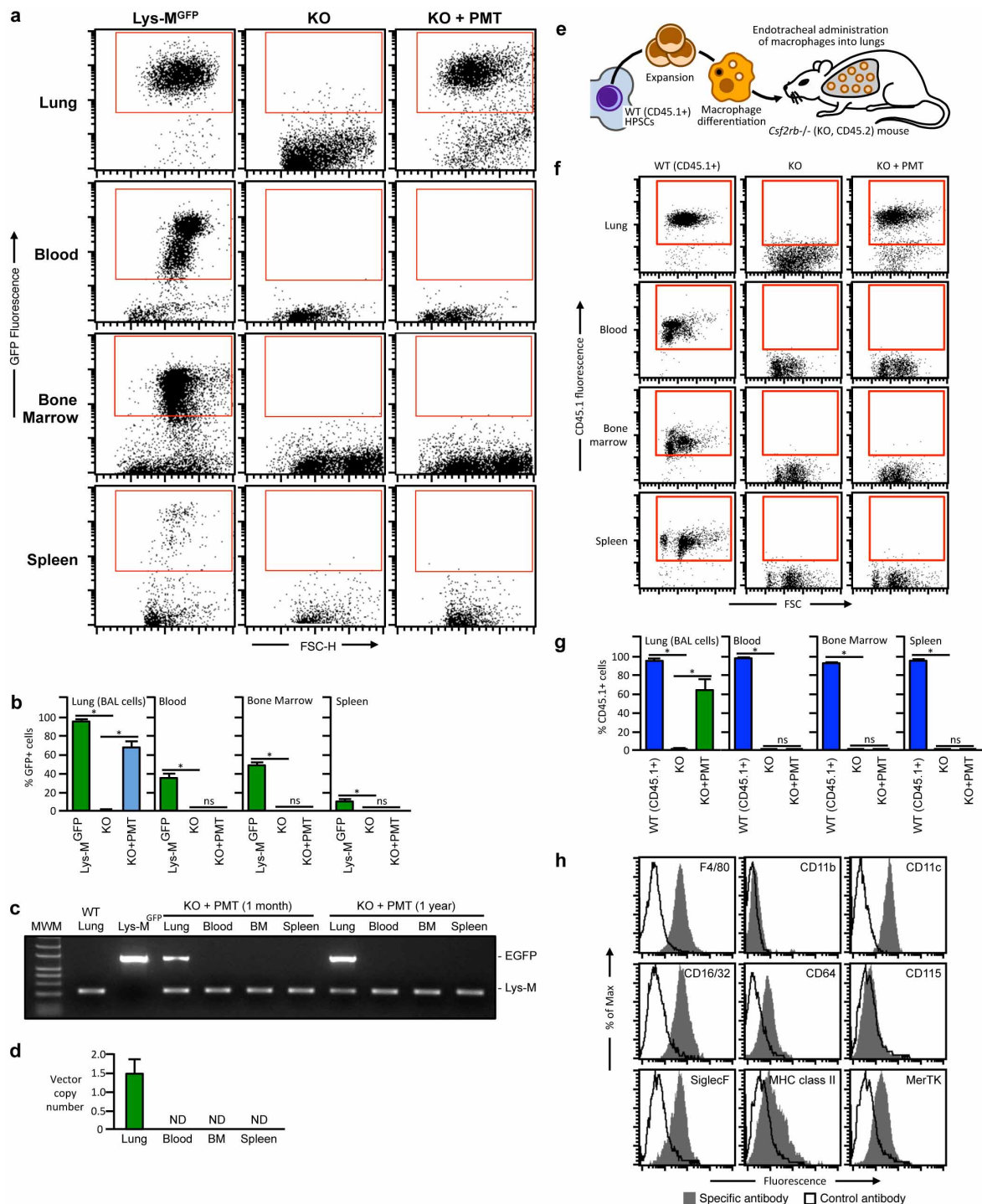
determinations per condition). **f**, **g**, Evaluation of surfactant clearance capacity. Representative photomicrographs of BMDMs from WT (left) or *Csf2rb*<sup>-/-</sup> (right) were examined before (top) or immediately after incubation with surfactant for 24 h (middle), or after exposure, removal of extracellular surfactant and culture for 24 h in the absence of surfactant (lower) after oil-red-O staining (representative of  $n = 3$  per condition). Scale bar, 20  $\mu\text{m}$ . **g**, Measurement of surfactant clearance by BMDMs after exposure as just described (**f**) and quantified using a visual grading scale (the oil-red-O staining index) to measure the degree of staining. Bars represent the mean  $\pm$  s.e.m. ( $n = 3$  per condition) of oil-red-O staining score for 10 high-power fields for each group. ND, not detected; ns, not significant; \*\*\* $P < 0.001$ .





**Extended Data Figure 3 | Efficacy of PMT in *Csf2rb*<sup>-/-</sup> mice and characterization of macrophages after PMT.** **a**, Detection of CD131 (top) or actin (bottom) in BAL cells by western blotting 1 year after PMT (each lane represents one mouse of 6 per group). **b**, Representative cytology of BAL obtained 1 year after PMT after staining with PAS or oil red O (ORO) (6 mice per group). Scale bar, 25  $\mu$ m. Oil-red-O positive cells were seen rarely in WT mice and occasionally in PMT-treated *Csf2rb*<sup>-/-</sup> mice (insets). Cytological abnormalities in BAL from untreated *Csf2rb*<sup>-/-</sup> mice including large, 'foamy', PAS- and oil-red-O-stained alveolar macrophages and PAS-stained cellular debris, were corrected by PMT. **c**, Representative photomicrographs of PAS-stained whole-mount lung sections 1 year after PMT. Note that some residual disease remained at 1 year (original magnification,  $\times 1$ ). **d**, GFP<sup>+</sup> cells in BAL cells from WT or *Csf2rb*<sup>-/-</sup> mice 2 months after PMT of Lys-M<sup>GFP</sup> BMDMs (representative of  $n = 3$  (WT) or  $n = 6$  (*Csf2rb*<sup>-/-</sup>) mice) (original magnification,  $\times 20$ ). **e**, Macrophage replication after PMT. *Csf2rb*<sup>-/-</sup> mice received Lys-M<sup>GFP</sup> BMDMs by PMT and paraffin-embedded lung was immunostained for Ki67 1 month or 1 year later. Scale bar, 50  $\mu$ m; inset, 10  $\mu$ m. **f**, Ki67 staining of BAL cells from untreated WT mice (**e**). Inset shows positive

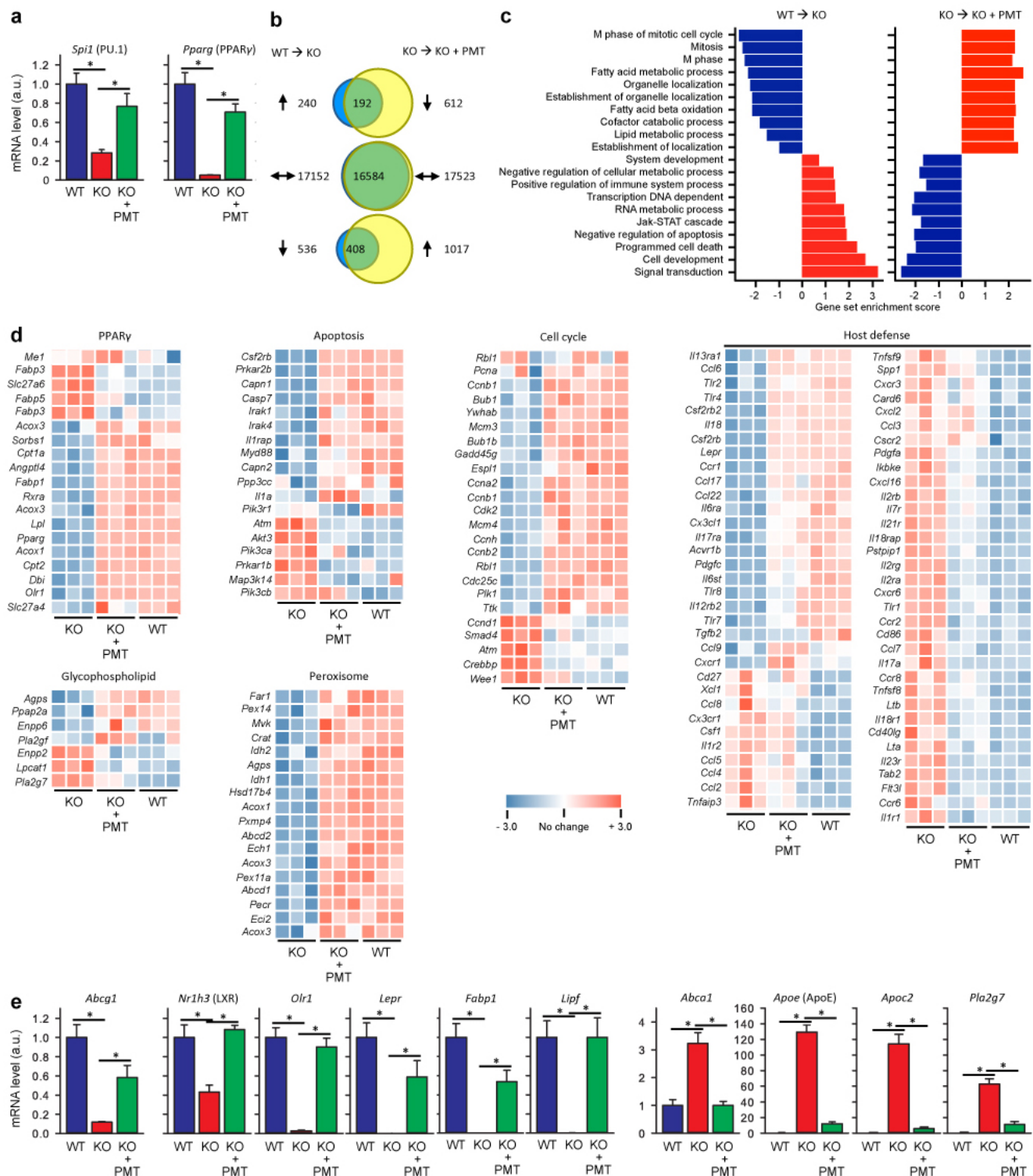
(left) or negative (right) staining. Scale bar, 50  $\mu$ m; inset 10  $\mu$ m. Graph shows the per cent Ki67<sup>+</sup> BAL cells in age-matched WT mice ( $n = 5$ ). **g**, Representative immunofluorescence photomicrographs of frozen lung sections 1 year after PMT of Lys-M<sup>GFP</sup> into *Csf2rb*<sup>-/-</sup> mice identifying GFP<sup>+</sup> cells (top), Ki67<sup>+</sup> cells (middle) and GFP<sup>+</sup> Ki67<sup>+</sup> (replicating, PMT-derived) cells (bottom) (representative of  $n = 3$  mice). Scale bar, 20  $\mu$ m; inset scale bar, 10  $\mu$ m. Quantitative summary data are shown in Fig. 2c. **h**, Localization of macrophages within the lungs 1 year after PMT of Lys-M<sup>GFP</sup> BMDMs into *Csf2rb*<sup>-/-</sup> mice and visualization in frozen lung sections after CD68 immunostaining, DAPI counter staining, and fluorescence microscopy to detect CD68<sup>+</sup> GFP<sup>+</sup> cells (that is, PMT-derived macrophages) or CD68<sup>+</sup> GFP<sup>-</sup> cells (that is, non-PMT-derived endogenous macrophages). Graph shows quantitative data for  $n = 6$  mice. **i**, Localization of macrophages in these same mice (**h**) by detecting GFP by immunohistochemical staining of paraffin-embedded lung sections using light microscopy to eliminate potential interference from autofluorescence (representative of  $n = 6$  mice). Quantitative summary data are shown in Fig. 3b.



**Extended Data Figure 4 | Tissue distribution and characterization of transplanted cells 1 year after PMT.** **a–d**, Two-month-old *Csf2rb*<sup>-/-</sup> mice (4 per group) received one PMT of Lys-M<sup>GFP</sup> BMDMs. Twelve months later, untreated, age-matched WT Lys-M<sup>GFP</sup> or *Csf2rb*<sup>-/-</sup> mice and PMT-treated *Csf2rb*<sup>-/-</sup> mice were evaluated using flow cytometry to detect GFP<sup>+</sup> cells in the indicated organs. Representative data (**a**) and the percentage of GFP<sup>+</sup> cells in the gated region are shown (**b**). Similar results were observed in *Csf2rb*<sup>-/-</sup> mice 2 months after PMT of Lys-M<sup>GFP</sup> BMDMs except the percentage of GFP<sup>+</sup> BAL lung cells was not quantified (not shown). **c**, Detection of Lys-M<sup>GFP</sup> PMT cells by PCR. PCR of genomic DNA from BAL cells (Lung), white blood cells (Blood), bone marrow (BM) cells and splenocytes (Spleen) 1 month or 1 year after Lys-M<sup>GFP</sup> BMDM PMT was performed to detect EGFP and Lysozyme M gene. BAL cells (Lung) from WT and Lys-M<sup>GFP</sup> were shown as negative and positive control for EGFP. EGFP was only detected in lung. **d**, Vector copy number analysis after gene-corrected BMDM PMT.

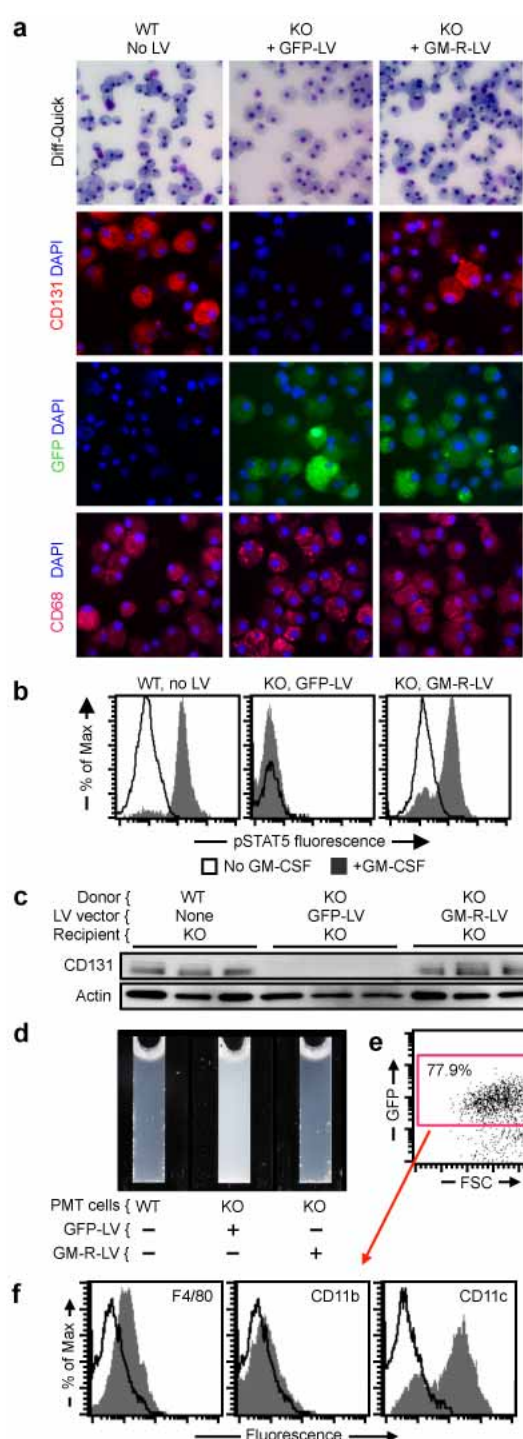
Quantitative PCR with vector-specific primers (R-U5) was performed using genomic DNA from BAL cells (Lung), white blood cells (Blood), bone marrow (BM) cells and splenocytes (Spleen) obtained 1 year after PMT of gene-corrected macrophages. Note that the viral vector was only detected in lung. **e–h**, CD45.2<sup>+</sup> *Csf2rb*<sup>-/-</sup> mice received one PMT of CD45.1<sup>+</sup> BMDMs from congenic WT mice (**e**) and 1 year later, untreated, age-matched WT (CD45.1<sup>+</sup>) or *Csf2rb*<sup>-/-</sup> (CD45.2<sup>+</sup>) mice and PMT-treated *Csf2rb*<sup>-/-</sup> mice were evaluated by flow cytometry to detect CD45.1<sup>+</sup> cells in the indicated organs. Representative data (**f**) and the percentage of CD45.1<sup>+</sup> cells in the indicated organs are shown (**g**). Phenotypic characterization of PMT-derived (CD45.1<sup>+</sup>) cells (as shown in the gated region (**f**)). Results are similar to those for PMT of Lys-M<sup>GFP</sup> BMDMs (Fig. 3d). Numeric data are mean  $\pm$  s.e.m. of  $n = 4$  mice per group (**b**, **d**) or  $n = 5$  mice per group (**g**). ND, not detected. \* $P < 0.05$ . ns, not significant.





**Extended Data Figure 5 | Global gene expression analysis of alveolar macrophages from age-matched WT, *Csf2rb*<sup>-/-</sup> and *Csf2rb*<sup>-/-</sup> mice 1 year after PMT of WT BMDMs. a**, Expression of *Spi1* (PU.1) and *Pparg* (PPAR $\gamma$ ) were confirmed by qRT-PCR using independent samples (6 mice per group). **b**, Venn diagrams showing numbers of genes whose expression was altered in alveolar macrophages from *Csf2rb*<sup>-/-</sup> compared to WT mice (WT $\rightarrow$ KO) or PMT-treated compared to untreated *Csf2rb*<sup>-/-</sup> mice (KO $\rightarrow$ KO+PMT). Only genes with statistically significant changes (false detection rate <10%) of at least twofold were marked as increased (up arrows) or decreased (down arrows). The numbers of genes for which expression was disrupted in *Csf2rb*<sup>-/-</sup> mice and normalized by PMT (or unchanged in both comparisons) is shown in the overlap regions. **c**, Gene ontology analysis

identifying pathways disrupted in *Csf2rb*<sup>-/-</sup> mice and restored by PMT. Data show the coordinate increases (red) or decreases (blue) in expression of genes in all gene sets significant at or below a false detection rate of 10% calculated by the Gene Set Test with correction for multiple testing. **d**, Heat maps showing differentially expressed genes in multiple KEGG pathways including PPAR $\gamma$ -regulated genes, glycophospholipid metabolism, peroxisome function apoptosis, cell cycle control, and immune host defence. Genes with increased or decreased transcript levels are shown by red and blue colours, respectively. **e**, Confirmation by qRT-PCR for selected genes important in lipid metabolism, using independent samples. Data are mean  $\pm$  s.e.m. (6 mice per group). \* $P$  < 0.05.



**Extended Data Figure 6 | Effects of PMT of gene-corrected macrophages on hPAP.** **a**, Macrophages derived from *Csf2rb*<sup>-/-</sup> LSK cells transduced with GM-R-LV or GFP-LV, or from non-transduced WT LSK cells (indicated) were examined by light microscopy after DiffQuick staining (top), or by immunofluorescence microscopy after staining with anti-CD131 (GM-CSF-R-β) and DAPI (upper middle), DAPI alone (lower middle), or anti-CD68 and DAPI (bottom). Images are representative of three experiments per condition. **b**, Evaluation of GM-CSF receptor signalling in the indicated cells (before PMT) by measurement of GM-CSF-stimulated STAT5 phosphorylation by flow cytometry. Representative of *n* = 3 experiments per condition. Quantitative summary data are shown in Fig. 5b. **c**, Western blotting to detect GM-CSF receptor-β (CD131) (top) or actin (bottom, as a loading control) in BAL cells from age-matched *Csf2rb*<sup>-/-</sup> mice 2 months after PMT as indicated (each lane represents one mouse of *n* = 10, 8, 10 per group, respectively). **d**, Appearance of BAL from age-matched *Csf2rb*<sup>-/-</sup> mice 2 months after PMT as indicated (representative of *n* = 10, 8, 10 per group, respectively). **e**, **f**, One year after PMT of GM-R-LV transduced *Csf2rb*<sup>-/-</sup> LSK cell-derived macrophages in *Csf2rb*<sup>-/-</sup> mice, GFP<sup>+</sup> cells were identified (**e**) and evaluated for cell surface markers by flow cytometry (**f**) (representative of *n* = 7 mice).



**Extended Data Table 1 | Oligonucleotide primers used to quantify mRNA transcripts by qRT-PCR and detection of PMT-derived cellular DNA by PCR**

TaqMan® Primers			
Gene name	Accession no.	Product (bp)	Catalogue no.
<i>Spi1</i> (PU.1)	NM_011355.1	91	Mm00488142_m1
<i>Pparg</i>	NM_001127330.1	101	Mm01184322_m1
<i>Abcg1</i>	NM_009593.2	65	Mm00437390_m1
<i>Csf1</i>	NM_001113529.1	70	Mm00432686_m1
<i>Csf2</i>	NM_009969.4	125	Mm01290062_m1
<i>Csf2rb</i>	NM_007780.4	60	Mm00655745_m1
<i>Nr1h3</i> (LXR)	NM_001177730.1	57	Mm00443451_m1
<i>Olr1</i>	NM_138648.2	64	Mm00454586_m1
<i>Lepr</i>	NM_001122899.1	97	Mm00440181_m1
<i>Fabp1</i>	NM_007980.2	116	Mm00433188_m1
<i>Lipf</i>	NM_026334.3	87	Mm00471152_m1
<i>Abca1</i>	NM_013454.3	55	Mm00442646_m1
<i>ApoE</i> (Apo E)	NM_009696.3	64	Mm01307192_m1
<i>Apoc2</i>	NM_009695.3	60	Mm00437571_m1
<i>Pla2g7</i>	NM_013737.5	111	Mm00479105_m1
<i>Gapdh</i>	NM_008084.2	107	4352932E
18S RNA	X03205.1	187	4310893E
Custom Primers			
Gene name	Accession no.	Product (bp)	Sequence (5' → 3')
<i>Lys-M<sup>GFP</sup></i>	NA - transgene	680	aag ctg ttg gga aag gag gg gtc gcc gat ggg ggt gtt ct
Lysozyme-M	M21049	220	aag ctg ttg gga aag gag gg tcg gcc agg ctg act cca ta

**Extended Data Table 2 | Effect of the number of macrophages transplanted on the efficacy of PMT therapy of hPAP in *Csf2rb*<sup>-/-</sup> mice**

Parameter	WT	KO	KO + PMT (Macrophages/dose x 10 <sup>6</sup> )			
			0.5	1	2	4
Turbidity O.D. 600 nm	0.0553 0.023-0.21	1.96 † 1.85-2.74	0.765 ‡ 0.599-0.823	0.685 ‡ 0.472-0.997	0.38 ‡ 0.283-0.685	0.536 ‡ 0.301-0.732
SP-D µg/ml BAL	75.9 51-84	2105 † 1739-2396	1475 ‡ 1367-2034	1414 ‡ 656-1951	911 ‡ 660-1179	1299 ‡ 762-1634
GM-CSF pg/ml BAL	0 0-0	40.8 † 21.4-54.5	28.1 ‡ 15.5-37.8	17.2 ‡ 12.2-20.5	13.8 ‡ 6.97-17.5	14.8 ‡ 10.4-18.0
M-CSF pg/ml BAL	0 0-0	45.0 † 32.3-81.9	30.4 § 14.2-36.0	25.4 § 14.8-36.3	21.7 § 20.1-42.7	29.3 § 23.5-40.5
MCP-1 pg/ml BAL	0.88 0-25.1	135 † 123-163	72.1 ‡ 36.9-121	57.5 ‡ 45.9-80.7	49.0 ‡ 26.1-63.0	64.4 ‡ 28.2-109
<i>Csf2rb</i> mRNA A.U.	1.05 0.86-1.08	0 † 0-0	0.108 ‡ 0.085-0.213	0.167 ‡ 0.095-0.82	0.265 ‡ 0.097-1.46	0.447 ‡ 0.197-0.88
<i>Spi1</i> mRNA A.U.	1.02 0.78-1.13	0.306 † 0.289-0.393	0.468 ‡ 0.27-0.470	0.474 ‡ 0.351-0.707	0.475 ‡ 0.367-0.803	0.480 ‡ 0.446-0.595
<i>Pparg</i> mRNA A.U.	0.929 0.902-1.13	0.052 † 0.0-0.106	0.241 ‡ 0.196-0.362	0.295 ‡ 0.267-0.607	0.516 ‡ 0.327-0.923	0.360 ‡ 0.318-0.603
<i>Abcg1</i> mRNA A.U.	0.933 0.833-1.12	0.08 † 0.07-0.153	0.148 ‡ 0.135-0.183	0.232 ‡ 0.134-0.327	0.179 ‡ 0.106-0.455	0.220 ‡ 0.173-0.229

A.U., arbitrary units; BAL, bronchoalveolar lavage; hPAP, hereditary pulmonary alveolar proteinosis; KO, *Csf2rb* knockout mice; O.D., optical density; PMT, pulmonary macrophage transplantation; WT, wild type.

\* Mice received the indicated numbers of WT BMDMs once by PMT. Three months later, BAL fluid and cells were obtained from PMT-treated knockout mice, and age-matched, untreated WT or knockout mice (7 mice per group for each condition evaluated). BAL turbidity, the concentration of SP-D, GM-CSF, M-CSF and MCP-1 in BAL fluid, and the relative abundance of *Csf2rb*, *Spi1* (P.U.1), *Pparg* and *Abcg1* mRNA transcripts in alveolar macrophages were measured as described in Methods. All data are presented as median (interquartile range (IQR)) and between-group comparisons were done using non-parametric methods for consistency since results for some groups were either undetectable, not normally distributed or of unequal variance.

† Result is significantly different compared to untreated WT mice (Mann-Whitney rank sum test,  $P < 0.001$ ).

‡ Result is significantly different compared to untreated KO mice (Kruskal-Wallis One Way Analysis of Variance on Ranks with Pairwise comparison to untreated KO mice by the Student-Neuman-Keuls method,  $P < 0.05$ ).

§ Result is not significantly different compared to untreated KO mice (Kruskal-Wallis One Way Analysis of Variance on Ranks,  $P = 0.133$ ).

**Extended Data Table 3** | Comparison of the effects of single versus repeated macrophage administrations on the efficacy of PMT therapy of hPAP in *Csf2rb*<sup>-/-</sup> mice

Parameter	Number of PMT Administrations		P-value †
	One	Four	
Turbidity, O.D. 600 nm	2.014 (1.77-2.53)	1.68 (1.49-3.29)	0.486
SP-D, µg/ml BAL	816 (750-996)	772 (653-796)	0.486
GM-CSF, pg/ml BAL	18.3 (15.3-35.6)	14.3 (13.8-27.8)	0.20
M-CSF, pg/ml BAL	55.5 (50.8-65.6)	29.7 (28.1-49.8)	0.114
MCP-1, pg/ml BAL	88.3 (74.0-118)	53.6 (35.2-78.8)	0.114
<i>Spi1</i> mRNA, A.U.	0.377 (0.284-0.545)	0.322 (0.268-0.362)	0.686
<i>Pparg</i> mRNA, A.U.	0.201 (0.122-0.474)	0.234 (0.169-0.303)	1.0
<i>Abcg1</i> mRNA, A.U.	0.116 (0.098-0.236)	0.117 (0.083-0.131)	0.886

A.U., arbitrary units; BAL, bronchoalveolar lavage; hPAP, hereditary pulmonary alveolar proteinosis; KO, *Csf2rb* knockout mice; O.D., optical density; PMT, pulmonary macrophage transplantation.

\* Knockout mice received  $2 \times 10^6$  macrophages by PMT either once or as four monthly doses. Four months after the initial PMT administration in both groups, BAL fluid and cells were obtained from all mice (6 per group for each condition evaluated). BAL turbidity, the concentration of SP-D, GM-CSF, M-CSF and MCP-1 in BAL fluid, and the relative abundance of *Spi1* (PU.1), *Pparg* and *Abcg1* mRNA transcripts in alveolar macrophages were measured as described in Methods. All data are presented as median (interquartile range (IQR)) and between-group comparisons were done using non-parametric methods for consistency since results for some groups were either not normally distributed or of unequal variance.

† Mann-Whitney Rank Sum Test. P-values of  $\leq 0.05$  was considered to indicate statistical significance.



**Extended Data Table 4 | Effect of PMT of WT or gene-corrected macrophages on haematological indices and lung proinflammatory cytokine levels**

Blood Safety Evaluation - PMT Using Wild-Type Macrophages				
Hematologic Parameter	Normal range	WT (n=6)	KO (n=6)	KO + PMT (n=6)
Hemoglobin, g/dL	11.0 – 15.1	12.7 (12.2 – 13.2)	14.5 (13.9 – 15.1) ‡	12.9 (12.2 – 13.4) ¶
Hematocrit, %	35.1 – 45.4	46.5 (44.5 – 47.3)	53.0 (50.2 – 56.3) ‡	47.2 (43.0 – 50.5) ¶
WBC, x10 <sup>3</sup> /μl	1.8 – 10.7	2.53 (1.53 – 5.35)	5.08 (3.93 – 6.34) †	3.27 (2.80 – 4.10) ¶
Neutrophils, x10 <sup>3</sup> /μl	0.1 – 2.4	0.29 (0.107 – 0.67)	0.89 (0.125 – 2.63) †	0.670 (0.102 – 1.41) §
Lymphocytes, x10 <sup>3</sup> /μl	0.9 – 9.3	2.18 (1.06 – 4.26)	3.24 (1.60 – 5.79) †	2.37 (1.76 – 2.94) §
Monocytes, x10 <sup>3</sup> /μl	0.0 – 0.4	0.15 (0.11 – 0.24)	0.155 (0.12 – 0.32) †	0.19 (0.15 – 0.25) §
Eosinophils, x10 <sup>3</sup> /μl	0.0 – 0.2	0.01 (0.01 – 0.04)	0.02 (0.01 – 0.06) †	0.015 (0.01 – 0.02) §
Basophils, x10 <sup>3</sup> /μl	0.0 – 0.2	0.0 (0.0 – 0.01)	0.01 (0.01 – 0.01) †	0.0 (0.0 – 0.003) ¶
Platelets, x10 <sup>3</sup> /μl	592 – 2972	558 (463 – 735)	1035 (830 – 1145) ‡	993 (838 – 1031) §
Blood Safety Evaluation - PMT Using Gene-Corrected Macrophages				
Hematologic Parameter	Normal range	WT (n=5)	KO (n=5)	KO + PMT (n=7)
Hemoglobin, g/dL	11.0 – 15.1	10.2 (8.5 – 10.8)	16.9 (13.8 – 18.5) ‡	12.2 (11.1 – 12.9) ¶
Hematocrit, %	35.1 – 45.4	37.5 (33.3 – 40.4)	68.8 (55.4 – 80.3) ‡	49.4 (48.8 – 55.9) ¶
WBC, x10 <sup>3</sup> /μl	1.8 – 10.7	1.90 (1.30 – 10.7)	6.48 (5.11 – 8.57) †	5.08 (2.32 – 5.88) §
Neutrophils, x10 <sup>3</sup> /μl	0.1 – 2.4	1.31 (0.67 – 8.01)	2.61 (1.71 – 2.85) †	1.49 (0.96 – 2.89) §
Lymphocytes, x10 <sup>3</sup> /μl	0.9 – 9.3	0.65 (0.26 – 1.38)	3.39 (2.70 – 5.15) ‡	2.03 (1.15 – 3.33) ¶
Monocytes, x10 <sup>3</sup> /μl	0.0 – 0.4	0.25 (0.15 – 0.90)	0.44 (0.29 – 0.87) †	0.25 (0.15 – 0.40) §
Eosinophils, x10 <sup>3</sup> /μl	0.0 – 0.2	0.01 (0.01 – 0.43)	0.02 (0.02 – 0.07) †	0.08 (0.01 – 0.16) §
Basophils, x10 <sup>3</sup> /μl	0.0 – 0.2	0 (0.0 – 0.07)	0.01 (0.00 – 0.04) †	0.01 (0.00 – 0.03) §
Platelets, x10 <sup>3</sup> /μl	592 – 2972	619 (441 – 1478)	1229 (1094 – 1460) †	1381 (872 – 1614) §
Lung Safety Evaluation - PMT Using WT Macrophages				
Cytokine in BAL Fluid		WT (n=6)	KO (n=6)	KO + PMT (n=6)
IL-6, pg/ml		3.01 (1.35 – 3.89)	164 (43.6 – 364) ‡	11.02 (4.87 – 60.1) ¶
IL-1β, pg/ml		0 (0.0 – 0.0)	3.49 (2.42 – 9.52) ‡	0.92 (0.04 – 1.02) ¶
TNFα, pg/ml		0.52 (0.0 – 0.94)	12.4 (7.22 – 17.9) ‡	2.75 (1.08 – 3.17) ¶
Lung Safety Evaluation - PMT Using Gene Corrected Macrophages				
Cytokine in BAL Fluid		WT (n=5)	KO (n=5)	KO + PMT (n=7)
IL-6, pg/ml		0.30 (0.0 – 2.43)	16.2 (14.1 – 62.1) ‡	29.5 (13.7 – 42.5) §
IL-1β, pg/ml		0 (0.0 – 0.29)	2.01 (0.47 – 8.1) †	1.65 (0.94 – 3.08) §
TNFα, pg/ml		2.45 (0.46 – 2.76)	6.14 (3.68 – 7.99) ‡	3.06 (2.45 – 7.98) §

BAL, bronchoalveolar lavage; hPAP, hereditary pulmonary alveolar proteinosis; KO, *Csf2rb* knockout mice; O.D., optical density; PMT, pulmonary macrophage transplantation; WT, wild type.

\* Knockout mice received WT or *Csf2rb* gene-corrected knockout macrophages ( $2 \times 10^6$  cells per mouse) once by PMT and 12 months later, blood and BAL fluid were obtained from PMT-treated knockout mice, or age-matched, untreated knockout or WT mice and evaluated as described in Methods. Number of mice per group is indicated. All data are presented as median (interquartile range (IQR)) and between-group comparisons were done using non-parametric methods (Mann-Whitney rank sum test) for consistency since results for some groups were either not normally distributed or of unequal variance. *P* values  $\leq 0.05$  were considered to be significant.

† Result is not significantly different compared to WT mice.

‡ Result is significantly different compared to WT mice.

§ Result is not significantly different compared to untreated knockout mice.

¶ Result is significantly different compared to untreated knockout mice.

# Structure and immune recognition of trimeric pre-fusion HIV-1 Env

Marie Pancera<sup>1</sup>, Tongqing Zhou<sup>1</sup>, Aliaksandr Druz<sup>1</sup>, Ivelin S. Georgiev<sup>1</sup>, Cinque Soto<sup>1</sup>, Jason Gorman<sup>1</sup>, Jinghe Huang<sup>2</sup>, Priyamvada Acharya<sup>1</sup>, Gwo-Yu Chuang<sup>1</sup>, Gilad Ofek<sup>1</sup>, Guillaume B. E. Stewart-Jones<sup>1</sup>, Jonathan Stuckey<sup>1</sup>, Robert T. Bailer<sup>1</sup>, M. Gordon Joyce<sup>1</sup>, Mark K. Louder<sup>1</sup>, Nancy Tumba<sup>3</sup>, Yongping Yang<sup>1</sup>, Baoshan Zhang<sup>1</sup>, Myron S. Cohen<sup>4</sup>, Barton F. Haynes<sup>5</sup>, John R. Mascola<sup>1</sup>, Lynn Morris<sup>3,6,7</sup>, James B. Munro<sup>8</sup>, Scott C. Blanchard<sup>9</sup>, Walther Mothes<sup>8</sup>, Mark Connors<sup>2</sup> & Peter D. Kwong<sup>1</sup>

**The human immunodeficiency virus type 1 (HIV-1) envelope (Env) spike, comprising three gp120 and three gp41 subunits, is a conformational machine that facilitates HIV-1 entry by rearranging from a mature unliganded state, through receptor-bound intermediates, to a post-fusion state. As the sole viral antigen on the HIV-1 virion surface, Env is both the target of neutralizing antibodies and a focus of vaccine efforts. Here we report the structure at 3.5 Å resolution for an HIV-1 Env trimer captured in a mature closed state by antibodies PGT122 and 35022. This structure reveals the pre-fusion conformation of gp41, indicates rearrangements needed for fusion activation, and defines parameters of immune evasion and immune recognition. Pre-fusion gp41 encircles amino- and carboxy-terminal strands of gp120 with four helices that form a membrane-proximal collar, fastened by insertion of a fusion peptide-proximal methionine into a gp41-tryptophan clasp. Spike rearrangements required for entry involve opening the clasp and expelling the termini. N-linked glycosylation and sequence-variable regions cover the pre-fusion closed spike; we used chronic cohorts to map the prevalence and location of effective HIV-1-neutralizing responses, which were distinguished by their recognition of N-linked glycan and tolerance for epitope-sequence variation.**

Over the last 50 years, more than 70 million people have been infected or killed by the human immunodeficiency virus type 1 (HIV-1)<sup>1</sup>. A dominant contributing factor has been the molecular trickery of the HIV-1 envelope (Env) spike, a type I fusion machine that facilitates virus entry into cells by interacting with host cellular receptors and fusing membranes of virus and host cell (reviewed in ref. 2). Despite its exposed position on the viral membrane and the generation of narrow-breadth neutralizing antibody responses throughout the course of HIV-1 infection, the evolving HIV-1 Env spike successfully evades most antibody-mediated neutralization<sup>3</sup>. This evasion is, to a large degree, responsible for the difficulty in developing an effective HIV-1 vaccine.

Initially synthesized as a gp160 precursor, which is cleaved into gp120 and gp41 subunits, the trimeric HIV-1 Env spike displays unusual post-translational processing, including the addition of 25–30 N-linked glycans per gp120–gp41 protomer<sup>4</sup>, tyrosine sulphation<sup>5</sup>, and slow signal peptide cleavage<sup>6</sup>. Env rearranges from a pre-fusion mature closed state that evades antibody recognition through intermediate open states that bind to receptors, CD4 and co-receptor (either CCR5 or CXCR4), to a post-fusion state (reviewed in ref. 2). Over the last 20 years, substantial atomic-level detail has been obtained on these states, including structures of receptor-bound gp120<sup>7</sup>, post-fusion gp41<sup>8,9</sup>, and most recently the trimeric arrangement of pre-fusion gp120 along with two gp41 helices, one of which was aligned in sequence<sup>10,11</sup>. The pre-fusion structure of gp41 has, however, resisted atomic-level analysis. Because the primary structural rearrangement driving membrane fusion is the gp41 transition from pre-fusion to post-fusion conformations, the lack of a pre-fusion gp41 structure has stymied attempts to provide a coherent picture of

the conformational rearrangements the spike undergoes to facilitate entry.

Here we use neutralizing antibodies PGT122<sup>12</sup> and 35022<sup>13</sup> to capture the HIV-1 spike in a pre-fusion mature closed state. We obtained crystals of the antigen-binding fragments (Fabs) of these two antibodies in complex with a soluble, cleaved Env trimer construct (BG505 SOSIP.664)<sup>14–16</sup> and determined its atomic-level structure. Examination of this structure in the context of previously determined gp120 and gp41 structures affords a mechanistic understanding of the conformational transitions the spike undergoes to facilitate virus entry. We delineated aggregate parameters of glycan shielding and genetic variation and used infected donor serum to determine where the immune system succeeds in recognizing the HIV-1 spike. Analysis of the pre-fusion HIV-1 Env structure and its conformational rearrangements, combined with an understanding of its evasion from and vulnerabilities to the immune system, reveal similarities to other type I viral fusion machines as well as features of recognition by the human immune system unique to this important vaccine target.

## Structure determination and overall structure

Atomic-level information for virtually all of the HIV-1 Env ectodomain in its pre-fusion conformation has been obtained from antibody-bound complexes (Extended Data Fig. 1a). The recently determined crystal structure<sup>10</sup> of a soluble cleaved HIV-1 Env based on the BG505 SOSIP.664 construct was no exception; in particular, while an artificial disulphide and other modifications of the SOSIP.664 construct were critical to production of a homogeneous, soluble, cleaved trimer<sup>17</sup>, antibody

<sup>1</sup>Vaccine Research Center, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA. <sup>2</sup>HIV-Specific Immunity Section, Laboratory of Immunoregulation, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA. <sup>3</sup>Center for HIV and STIs, National Institute for Communicable Diseases of the National Health Laboratory Service (NHLS), Sandringham, Johannesburg 2131, South Africa. <sup>4</sup>Departments of Medicine, Epidemiology, Microbiology and Immunology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. <sup>5</sup>Duke University Human Vaccine Institute, Departments of Medicine, Surgery, Pediatrics and Immunology, Duke University School of Medicine, and the Center for HIV/AIDS Vaccine Immunology-Immunogen Discovery at Duke University, Durham, North Carolina 27710, USA. <sup>6</sup>University of the Witwatersrand, Braamfontein, Johannesburg 2000, South Africa. <sup>7</sup>Centre for the AIDS Programme of Research in South Africa (CAPRISA), University of KwaZulu-Natal, Durban 4041, South Africa. <sup>8</sup>Department of Microbial Pathogenesis, Yale University School of Medicine, New Haven, Connecticut 06536, USA. <sup>9</sup>Department of Physiology and Biophysics, Weill Cornell Medical College of Cornell University, New York, New York 10021, USA.

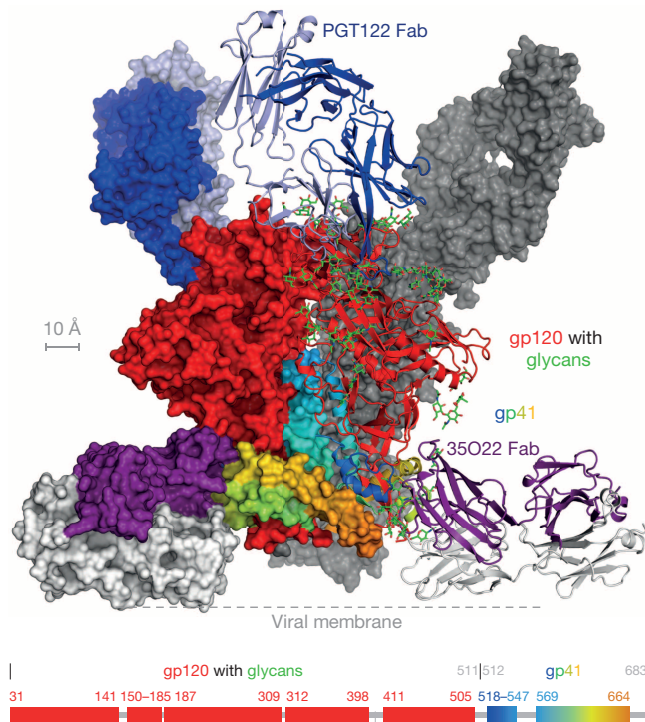
PGT122 appeared to facilitate crystallization<sup>10</sup>. Diffraction from crystals of the PGT122 complex, however, extended to only 4.7 Å resolution, hampering the trace of non-helical regions of gp41 as well as the placement and registry of side chains<sup>10</sup>. To obtain improved crystals, we explored the addition of antibody 35O22, which recognizes a gp120–gp41 epitope<sup>13</sup>. Addition of 35O22 to PGT122-bound viral spike in the membrane-bound virion context showed single-molecule fluorescence resonance energy transfer (smFRET) responses that closely resembled those of the mature native unliganded spike (Extended Data Fig. 1b)<sup>18</sup>. In the context of crystallization, addition of 35O22 to the PGT122–BG505 SOSIP.664 complex led to ternary complex crystals in space group P6<sub>3</sub>. Although diffraction was anisotropic, we succeeded in collecting ~3.5 Å data from a single crystal (Extended Data Table 1). Structure solution by molecular replacement with free structures of Fab PGT122<sup>19</sup>, Fab 35O22<sup>13</sup> and gp120<sup>20</sup> revealed a double antibody-bound gp120–gp41 protomer to occupy the asymmetric unit and led to an  $R_{\text{work}}/R_{\text{free}}$  of 21.35%/24.80%.

Overall, the HIV-1 spike forms a three-blade propeller, capped at its membrane-distal apex by antibody PGT122 and at the membrane-proximal end by antibody 35O22 (Fig. 1 and Extended Data Fig. 2a, b). Protomer interactions occur through assembled variable regions, V1, V2 and V3, which comprise the trimer association domain<sup>21</sup> at the membrane-distal portion of the spike, and also through gp41, primarily between helical interactions around the trimer axis<sup>10,11</sup>. No trimeric interactions are contributed by the gp120 core; indeed, a cleft or opening is found under the trimer association domains along the threefold axis where such associations might occur. Trimeric pre-fusion gp41 forms a platform through which the gp120 termini extend towards the viral

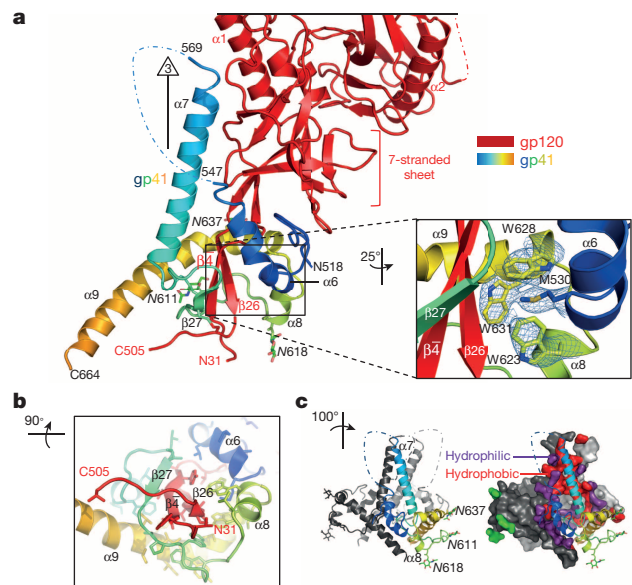
membrane. Unusually slow signal peptide cleavage<sup>6</sup>, which keeps the N terminus of gp120 proximal to the membrane, may facilitate folding of pre-fusion HIV-1 Env.

## Pre-fusion structure of gp41

Pre-fusion gp41 wraps its hydrophobic core around extended N and C termini-strands of gp120 (Fig. 2a). It forms a four-helix collar comprising helices  $\alpha 6$  (Met 530<sub>gp41</sub>–Asn 543<sub>gp41</sub>),  $\alpha 7$  (Gly 572<sub>gp41</sub>–Ile 595<sub>gp41</sub>),  $\alpha 8$  (Leu 619<sub>gp41</sub>–Trp 623<sub>gp41</sub>), and  $\alpha 9$  (Trp 628<sub>gp41</sub>–Asp 664<sub>gp41</sub>) (the numbering of pre-fusion gp41 helices and strands continues the nomenclature established for the gp120 subunit, which ends with helix  $\alpha 5$  and strand  $\beta 26$ ; for clarity, the molecule is named after each residue number). The first residue of gp41 visible in electron density corresponds to Val 518<sub>gp41</sub>, in the fusion peptide. An extended stretch connects to Leu 523<sub>gp41</sub>, which interacts hydrophobically with Trp 45<sub>gp120</sub> and Ile 84<sub>gp120</sub>, both of which are part of the seven-stranded  $\beta$ -sandwich around which the gp120 inner domain is organized<sup>22,23</sup>. The main chain of gp41 follows gp120 strand  $\beta 0$  away from the trimer axis towards the viral membrane until residue Met 530<sub>gp41</sub>, where the fold reverses itself and extends through  $\alpha 6$  towards the trimer axis and away from the viral membrane. Density between residues 547<sub>gp41</sub> and 569<sub>gp41</sub> is sparse (Extended Data Fig. 3a, b), and ultimately connects to helix  $\alpha 7$ , which forms a parallel coiled-coil about the trimer axis. At the end of  $\alpha 7$  is the gp41 cysteine loop (spanned by the Cys 598<sub>gp41</sub>–Cys 604<sub>gp41</sub> disulphide), whose C-terminal residues initiate strand  $\beta 27$  (Leu 602<sub>gp41</sub>–Thr 606<sub>gp41</sub>), which forms hydrogen bonds in an anti-parallel fashion with strand  $\beta 4$  ( $\beta$ -strand negative 4) from the N terminus of gp120. The intersubunit



**Figure 1 | Structure of a pre-fusion HIV-1 Env trimer bound by PGT122 and 35O22 antibodies.** One protomer and associated Fabs are shown in ribbon and stick representation, a second protomer in surface representation, and the third protomer in grey. Residues comprising the refined HIV-1 Env model are displayed on the bar, with beginning and final ordered residue of each segment labelled; vertical lines demark termini of the mature ectodomain subunits; unmodelled regions, which show residues disordered or not present in the BG505 SOSIP.664 construct, as well as glycans, which are disordered or not present, are shown in grey. 35O22 and PGT122 interactions with the HIV-1 Env trimer are shown in Extended Data Fig. 9a–f, and bound versus unbound Fabs are shown in Extended Data Fig. 9g.



**Figure 2 | Pre-fusion structure of gp41.** **a**, gp41 forms a four-helix collar, which wraps around extended N and C termini of gp120. Both gp120 (red) and gp41 (rainbow from blue to orange) are depicted in ribbon representation, with select residues and secondary structure labelled (additional labels are shown in Extended Data Fig. 10). The location of the trimer axis is indicated with a '3' inside a triangle. The orientation shown here is similar to that of Fig. 1, with perpendicular orientations provided in **b** and **c**. Zoom insert: the gp41 collar is clasped by the insertion of Met 530<sub>gp41</sub> into a tryptophan sandwich and by the alignment of helices  $\alpha 6$  and  $\alpha 8$ .  $2F_o - F_c$  electron density for clasp residues is depicted at  $1\sigma$ . **b**, gp41 holds the N and C termini of gp120 in its hydrophobic core. Colouring and representation are the same as in **a**, except that hydrophobic side chains are shown in stick representation and the orientation is rotated 90°, to depict the view from the viral membrane. **c**, gp41–trimer interfaces as viewed from side in ribbon and surface representation. Overall, the pre-fusion structure of gp41 and its trimeric arrangement appear to have no close structural relatives in the Protein Data Bank (Supplementary Table 2). N618, N611 and N637 indicate gp41 residues with attached N-linked glycans (see Fig. 1 bar).



disulphide ('SOS')<sup>14</sup> between residues 501<sub>gp120</sub> and 605<sub>gp41</sub> welds the C terminus of gp120 to the membrane-proximal end of strand  $\beta$ 4 (Fig. 2a). Upon passing the gp120 termini, gp41 reaches  $\alpha$ 8, whose C terminus aligns spatially with the N terminus of  $\alpha$ 6. After  $\alpha$ 8, the  $\alpha$ 9 helix reverses direction, again wrapping past the N and C termini of gp120, before extending horizontally along the edge of the spike to reach the gp120 termini from a neighbouring protomer.

Topologically, the gp41 subunit completes a single circle around the gp120 termini with the insertion of a hydrophobic prong comprising the side chain of Met 530<sub>gp41</sub> (which is located at the N terminus of  $\alpha$ 6, proximal to the fusion peptide), into a triple-tryptophan clasp formed by Trp 623<sub>gp41</sub> (from the C terminus of  $\alpha$ 8), Trp 628<sub>gp41</sub> (from the N terminus of  $\alpha$ 9) and Trp 631<sub>gp41</sub> (one turn into  $\alpha$ 9) (Fig. 2a insert). The alignment of helices  $\alpha$ 6 and  $\alpha$ 8 provides electrostatic complementarity that helps to stabilize the neighbouring methionine–tryptophan clasp.

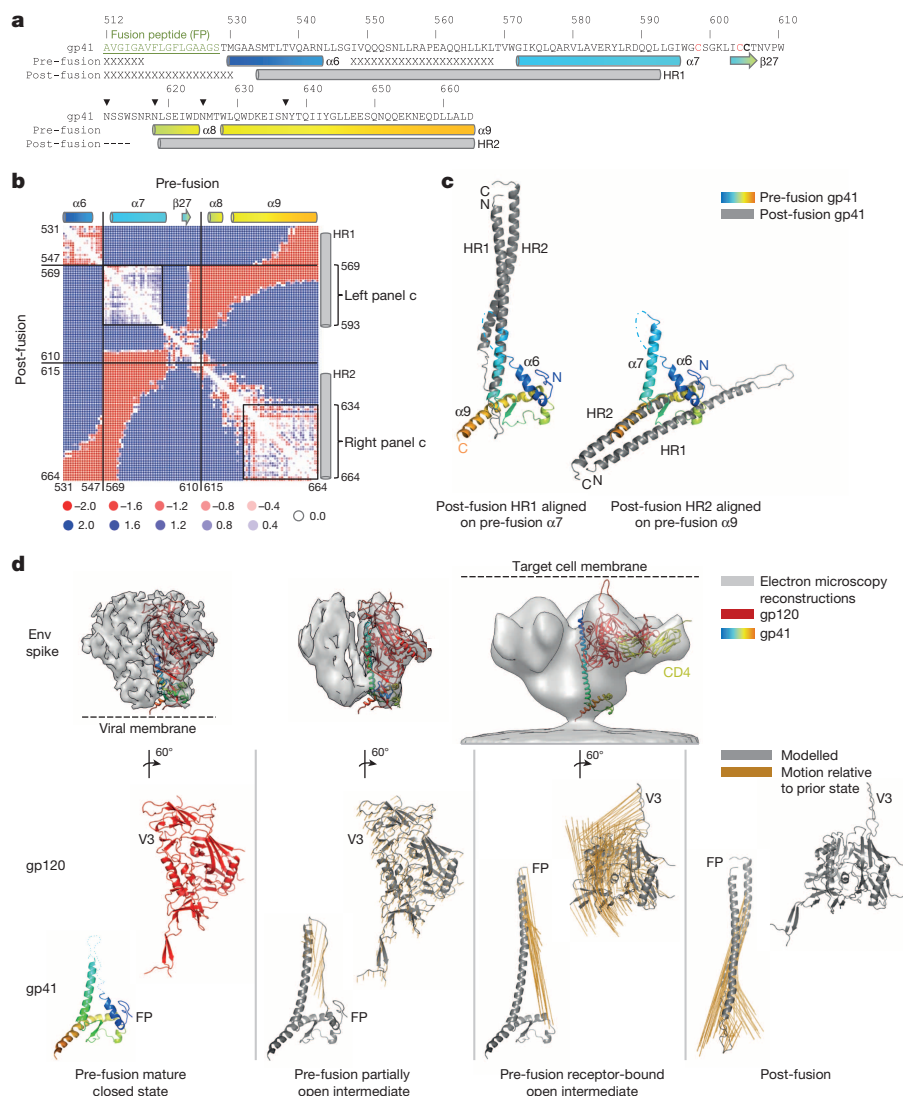
Within a single protomer, the buried surface area between gp41 and gp120 totals 5,270 Å<sup>2</sup>, including 216 Å<sup>2</sup> from glycan–protein interactions (Supplementary Table 1). A substantial portion of this is hydrophobic: gp41 essentially wraps its hydrophobic core around the N and C termini of gp120 (Fig. 2b). Trimer interfaces also bury a large surface area (3,140 Å<sup>2</sup> contributed by each protomer, comprising 1,920 Å<sup>2</sup> from the gp41–gp41 interface, 861 Å<sup>2</sup> from the gp120–gp120 interface and 360 Å<sup>2</sup> from the gp120–gp41 interface) (Extended Data Fig. 2c–f). Close to the trimer axis, these involve helix  $\alpha$ 7, as well as the N-terminal portion of the gp41–cysteine loop. Further from the trimer axis, interactions

involve  $\alpha$ 9. Other than interactions of  $\alpha$ 7, most inter-protomer interactions are hydrophilic (Fig. 2c).

### Pre-fusion to post-fusion gp41 transition

To understand the conformational transition from pre-fusion to post-fusion gp41, we compared the gp41 pre-fusion structure in our antibody-bound HIV-1 Env trimer with previously determined post-fusion structures<sup>8,9,24,25</sup> (Fig. 3). Post-fusion gp41 comprises two helices, HR1 and HR2 (Fig. 3a); these form a trimeric six-helical bundle, with HR1 helices arranged as an interior parallel coiled-coil, and exterior HR2 helices packed in an anti-parallel fashion to bring N-terminal fusion peptides and C-terminal transmembrane regions into proximity. Distance difference analysis<sup>26</sup> (Fig. 3b) of pre-fusion and post-fusion structures indicated two regions of structural similarity, corresponding to the pre-fusion  $\alpha$ 7 helix aligned with the C-terminal half of the post-fusion HR1 helix (Fig. 3c, left) and the pre-fusion  $\alpha$ 9 helix aligned with much of the post-fusion HR2 helix (Fig. 3c, right).

Superposition of pre-fusion  $\alpha$ 7 and post-fusion HR1 placed residues 569<sub>gp41</sub>–593<sub>gp41</sub> within 5 Å, with a root-mean-square deviation (r.m.s.d.) of 1.35 Å (Fig. 3c, left). For this superposition to occur, C $\alpha$ -movements of over 80 Å are required for the gp41 fusion peptide and  $\alpha$ 6 helix as well as for the C-terminal portion of the  $\alpha$ 9 helix. Notably, this superposition preserves the coiled-coil trimeric interactions of both pre-fusion and post-fusion molecules and thus is likely to mimic the natural conformational transition that occurs during membrane fusion. Meanwhile,



**Figure 3 | Entry rearrangements of HIV-1 Env.** **a**, BG505 sequence<sup>46</sup> of gp41, with pre-fusion and post-fusion secondary structure. Fusion peptide (FP) is underlined and labelled green. Several post-fusion gp41 structures have been determined ranging from a minimal, protease-treated, crystal structure (residues 556<sub>gp41</sub>–581<sub>gp41</sub>; 628<sub>gp41</sub>–661<sub>gp41</sub>; PDB ID 1AIK<sup>8</sup>) with 80% sequence identity to BG505<sup>46</sup> to a more complete gp41 structure (residues 531<sub>gp41</sub>–581<sub>gp41</sub>; 624<sub>gp41</sub>–681<sub>gp41</sub>; PDB ID 2X7R<sup>24</sup>) and an NMR structure that includes the cysteine loop (residues 539<sub>gp41</sub>–665<sub>gp41</sub>; PDB ID 2EZO<sup>25</sup>) of the simian immunodeficiency virus (SIV), which shares 48% sequence identity with BG505<sup>46</sup> and is substantially similar to the HIV-1 structures (less than 1 Å C $\alpha$  r.m.s.d. between overlapping residues of 1AIK and 2EZO). The post-fusion structure used here for comparisons was constructed from a chimera of HIV-1–SIV structures (Extended Data Fig. 3c). **b**, Difference distance analysis<sup>26</sup> of pre-fusion BG505 and post-fusion HIV-1–SIV chimeraic gp41. Secondary structure is indicated, along with missing residues of BG505 (548–568) and of SIV (611–614). **c**, Superposition of post-fusion gp41 (grey) onto pre-fusion gp41 (rainbow) for  $\alpha$ 7 (left) and  $\alpha$ 9 (right) pre-fusion helices. **d**, HIV-1 Env entry rearrangements. Electron microscopy reconstructions (top row) with gp120 (middle) and gp41 (bottom) rearrangements between each conformational state highlighted with orange lines depicting movement of each C $\alpha$  between conformations. Subunit models are shown in grey with modelling parameters and references provided in Extended Data Table 2. Antigenic recognition of each of these states is shown in Extended Data Fig. 5.

superposition of pre-fusion  $\alpha 9$  and post-fusion HR2 placed residues 634<sub>gp41</sub>–664<sub>gp41</sub> within 5 Å, with an r.m.s.d. of 3.58 Å (Fig. 3c, right); the substantial alignment of  $\alpha 9$  and HR2 helices indicates that the HR2 helix is preformed in the pre-fusion structure.

### Entry rearrangements of HIV-1 Env

Biosynthesis of HIV-1 Env starts with an uncleaved gp160 trimer. After cleavage, the spike condenses into the pre-fusion mature closed structure described here. In the gp120 inner domain, helix  $\alpha 1$  is formed, and a parallel strand exists between strands  $\beta 3$  and  $\beta 21$ ; in gp41, we observe helix  $\alpha 7$  to begin around residue 571<sub>gp41</sub>. A partially open electron microscopy structure<sup>27</sup> has been reported at 6 Å, in which the trimer association domains appear to be displaced from the trimeric axis, and helical density suggests helix  $\alpha 7$  to start several turns earlier; we modelled these rearrangements with a rigid body motion of 6 degrees for gp120 and the conversion of ~15 residues of helix  $\alpha 6$  and connecting stretch into helix  $\alpha 7$ , which extends ~20 Å towards the target cell membrane (Fig. 3d, middle panel; Extended Data Table 2).

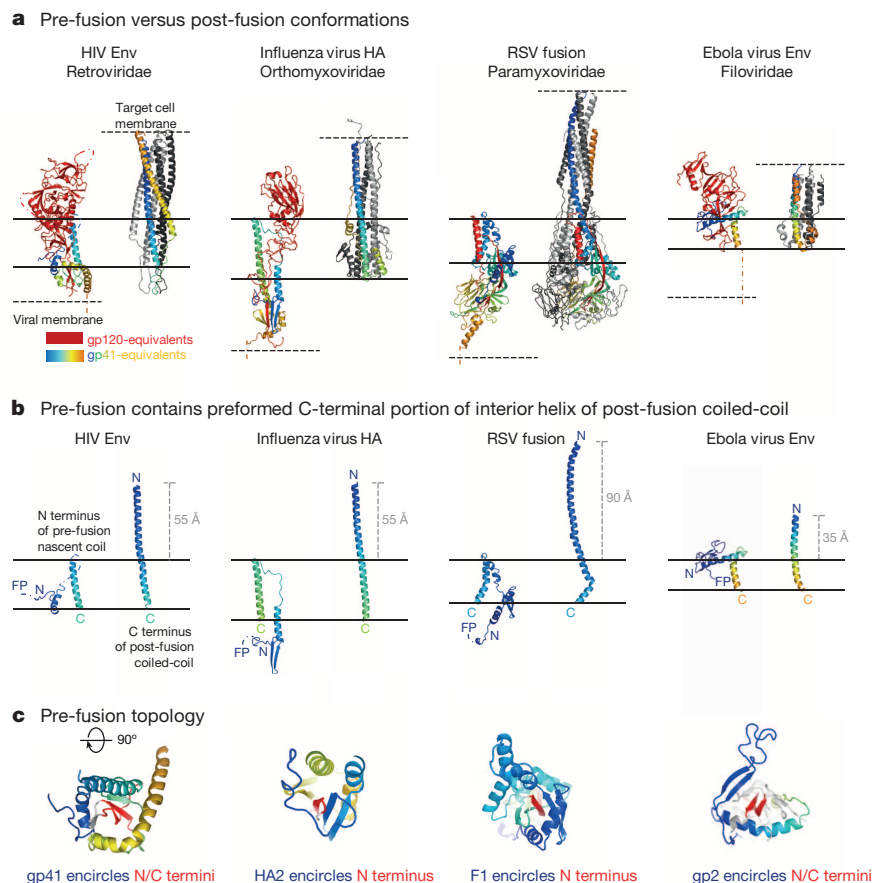
The CD4-bound state has been visualized by a number of electron microscopy reconstructions<sup>28,29</sup> and atomic-level structures<sup>7,22</sup>. In this state, V1V2 separates from V3: V3 points towards the target cell<sup>30</sup>, and the bridging sheet<sup>7</sup> assembles with strand  $\beta 2$  forming antiparallel hydrogen bonds with  $\beta 21$  (as opposed to the parallel  $\beta 3$ – $\beta 21$  interaction of the pre-fusion mature closed state; notably, the only parallel  $\beta$ -strand in the respiratory syncytial virus (RSV) F glycoprotein pre-fusion structure also changes conformation in RSV F pre- to post-fusion transition<sup>31</sup>). With layer 1 of the inner domain<sup>23</sup>, helix  $\alpha 0$  forms, and Gln428<sub>gp120</sub> and strand  $\beta 21$  invert; in layer 2, inner domain rearrangements include the swapping of distinct perpendicular interactions of Trp 112<sub>gp120</sub> and Trp 427<sub>gp120</sub> (Extended Data Fig. 4). CD4 binding allows HR2-peptide analogues (such as C34) to bind<sup>32</sup>, and we can model helix  $\alpha 7$  starting

as early as 554<sub>gp41</sub> with Met 530<sub>gp41</sub> still in its membrane-proximal tryptophan clasp, as expected because 35O22 binds the CD4-bound SOSIP. 664 (Extended Data Figs 3d, e and 5c, e). We envision that Env–CCR5 interactions<sup>33</sup> bring the CD4-bound state close to the target cell membrane, where ‘disassembling  $\alpha 6$ /assembling  $\alpha 7$  helices’ coupled to release of the Met 530<sub>gp41</sub> prong from its tryptophan clasp ultimately amasses the gp41 fusion peptide(s) (Fig. 3d, second panel from right, Extended Data Fig. 3f).

At this receptor-bound stage, it is easy to imagine the fusion peptide penetrating the target cell membrane, while strand  $\beta 27$  of the gp41-cysteine loop remains hydrogen-bonded to the gp120 termini (and the C terminus of the gp41 ectodomain remains in the viral membrane). Rearrangement of gp41 to its post-fusion conformation may be triggered by gp120 shedding<sup>34</sup>, with expulsion of the gp120 termini tugging on the gp41-cysteine loop and destabilizing pre-fusion gp41.

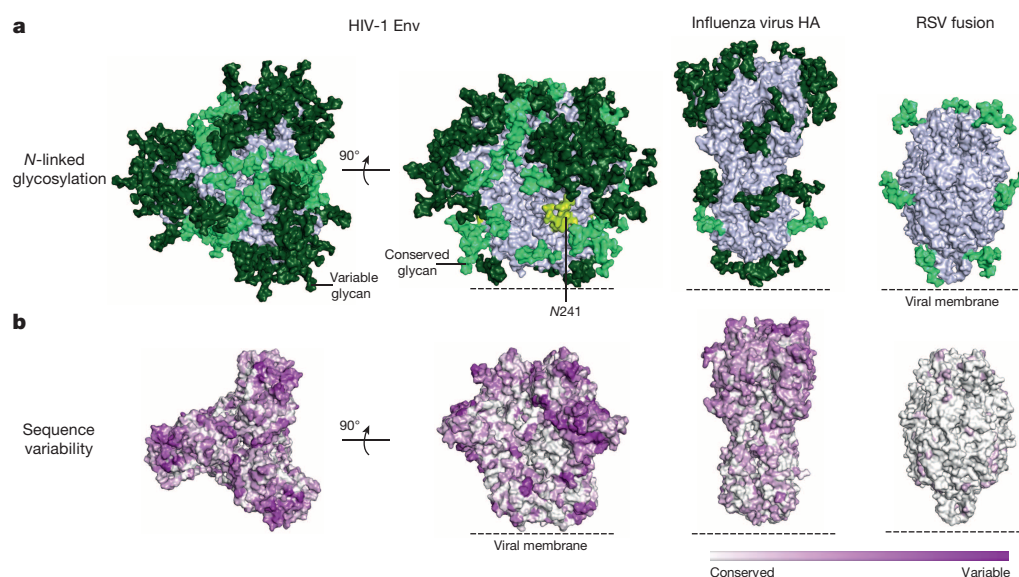
### HIV-1 rearrangements and other type I fusion machines

To determine whether the distinct elements we observed in pre-fusion gp41 were preserved elsewhere, we examined pre-fusion and post-fusion states of other type I fusion machines from influenza virus<sup>35,36</sup> (a member of the Orthomyxoviridae family of viruses), RSV<sup>31,37</sup> (Paramyxoviridae), and Ebola virus<sup>38,39</sup> (Filoviridae) (Fig. 4a). In all cases, a helix was observed in the gp41 pre-fusion equivalents, which corresponds in sequence to the C-terminal portion of the helix that in the post-fusion conformation comprises the interior coiled-coil characteristic of type I fusion machines<sup>8,9</sup> (Fig. 4b). With pre-fusion machines from HIV-1, influenza and Ebola viruses, the nascent pre-fusion helix adopts a coiled-coil; with RSV, a coiled-coil assembles immediately N-terminal to the nascent post-fusion helix. Despite marked differences in gp120-equivalents, similarity was observed in the overall topology of subunit interactions. Notably, all of the gp41-equivalents wrapped hydrophobic residues



**Figure 4 | Pre-fusion HIV-1 gp120-gp41 structure shares conserved structural and topological features with other type I fusion machines.** **a**, Pre-fusion (left) and post-fusion (right) structures. The pre-fusion structures are shown for a single protomer in ribbon representation with gp120-equivalent subunits in red, and gp41-equivalent subunits in rainbow (blue to orange). The trimeric post-fusion structures are shown with one subunit in rainbow (blue to orange), and the other in light and dark grey. **b**, The C-terminal portion of the preformed interior helix of post-fusion coiled-coil from **a** is shown, with fusion peptides (FP) and N- and C-terminal residues of post-fusion coiled-coils labelled, and the distance the inner coiled-coil extends between pre-fusion and post-fusion conformations indicated. **c**, The gp41 equivalents encircle extended  $\beta$ -strands of their gp120-equivalent partners. Ribbon representations are shown looking towards the viral membrane. With influenza virus, it is only the N terminus of the gp120 equivalent (HA1) that is wrapped by the gp41 equivalent (HA2), with the N terminus of HA2 completing about 20% more than a single encirclement. With RSV, it is also only the N terminus of the gp120 equivalent (F2) that is wrapped by the gp41 equivalent (F1), and the termini do not have to be expelled to transition to the post-fusion form. With Ebola, the gp41 equivalent (gp2) wraps both N and C termini strands of the gp120 equivalent (gp1), completing about 70% of a single encirclement. Such encirclement probably helps capture the energy of pre-fusion folding, which is released during the post-fusion transition to power membrane fusion.





**Figure 5 | Fully assembled shield revealed by pre-fusion HIV-1 gp120-gp41 trimer.** **a**, Glycan shield. Env N-linked glycans are depicted in light green (conserved; greater than 90% conservation) or dark green (variable; less than 90% conservation) on the pre-fusion mature closed Env structures for BG505 strain of HIV-1 (left), influenza virus H3 haemagglutinin (HA) (PDB ID 2YP7) (middle), and RSV fusion glycoprotein subtype A (PDB ID 4JHW) (right). A conserved glycan at residue 241<sub>gp120</sub> not present in the BG505 sequence is shown in yellow-green. **b**, Sequence variability.

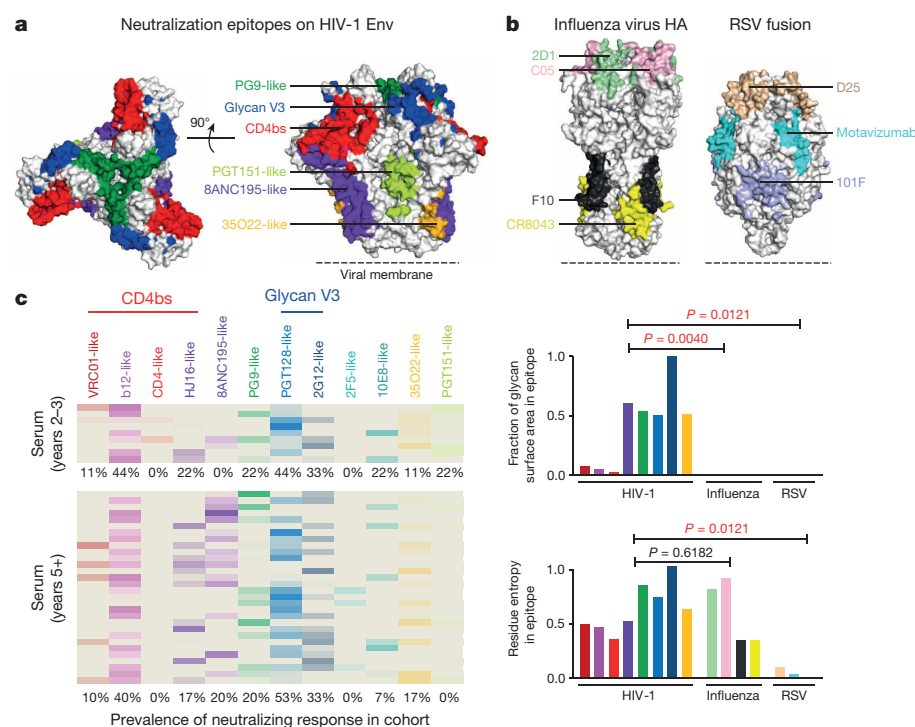
around extended termini (or N terminus) of their gp120-equivalents (Fig. 4c). Overall, the similarities in pre-fusion folding topology and in pre-fusion interior helices observed here, along with the previously observed similarity in post-fusion coiled-coils (reviewed in ref. 40), provide a more general and integrated view of the structural and conformational requirements of type I-mediated membrane fusion.

### Glycan shielding and genetic variation

The pre-fusion mature closed conformation of HIV-1 Env is the target of most neutralizing antibodies. The newly revealed structure of a near-complete gp120-gp41 Env trimer ectodomain provides an opportunity to understand aggregate properties of glycosylation and variation. Glycan shielding and genetic variation have long been recognized as mechanisms to avoid antibody recognition<sup>41</sup>. The BG505 SOSIP.664 sequence contains 28 sequons specifying N-linked glycosylation (including a T332N mutation). We modelled high mannose glycans (either Man5 or Man9) on each sequon and calculated accessible surface for radii ranging from

1.4 Å (the radius of a water molecule) to 10 Å (the approximate radius of a single immunoglobulin domain) (Extended Data Fig. 6). In the Man9-glycosylated model, 29% of the protein surface was solvent accessible, whereas only 3% of the surface was immunoglobulin-domain accessible. By contrast, with the fusion glycoproteins from influenza virus and RSV, 14% and 48%, respectively, of these surfaces were immunoglobulin-domain accessible (Fig. 5a).

In terms of genetic variation, we calculated the per-residue Shannon entropy of 3,943 sequences of HIV-1 (Fig. 5b). Approximately 50% of the surface was shown to have a variability of greater than 10%, a degree of surface variation shared by influenza virus, but not by RSV. When we combined glycan shielding and genetic variation, only ~2% of the surface was immunoglobulin accessible with a variability of less than 10% (Extended Data Fig. 7, upper panels); much of this conserved surface occurred at the membrane-proximal 'base' of the spike, which is expected to be sterically occluded by the viral membrane. To determine how this fully assembled shield compared to other conformations, we also assessed



**Figure 6 | Location and prevalence on the HIV-1 Env spike of neutralizing responses identified serologically from cohorts, 2-3 and 5+ years post-infection.** **a**, The location of the neutralization epitopes for broadly neutralizing antibodies is depicted on the pre-fusion mature closed Env spike with red for CD4-binding-site-directed antibody specificities (VRC01-, b12-, CD4-, and HJ16-like), purple for 8ANC195-like, green for V1V2-directed (PG9-like), blue for glycan-V3 specificities (PGT128- and 2G12-like), orange for 35O22-like specificities, and green-yellow for PGT151-like specificities. **b**, Top, broadly neutralizing epitopes on influenza virus HA (left, PDB ID 2YP7) and RSV fusion glycoprotein (right, PDB ID 4JHW). Bottom, glycan surface area and residue entropy of antibody epitopes for HIV-1, influenza virus and RSV, with bars coloured according to epitopes shown in **a** and **b** (except for epitopes not present in SOSIP.664 or where there is no atomic level definition). **c**, Neutralization fingerprint. For each serum, the predicted neutralization prevalence for each of the 12 antibody specificities is shown based on neutralization of 21 diverse HIV-1 strains (Extended Data Fig. 8).



the immunoglobulin accessibility of the CD4-bound conformation (Extended Data Fig. 7, lower panels). Notably, the CD4-bound conformation showed substantially higher levels of glycan-free, conserved surface, consistent with the greater ease by which antibodies reactive with the CD4-bound conformation are elicited—and by contrast, the difficulty in eliciting broadly neutralizing antibodies against the glycan-covered, sequence-variable pre-fusion closed state.

### Serologic recognition of pre-fusion closed Env

Despite multiple mechanisms of immune evasion that shield mature HIV-1 Env, potent broadly neutralizing antibodies do develop<sup>42</sup>. The structure of HIV-1 Env in the pre-fusion mature closed state allows us to map known epitopes on their most likely functional target (Fig. 6a) and to compare the recognition of broadly neutralizing HIV-1 antibodies, with those capable of neutralizing influenza virus and RSV (Fig. 6b).

To determine the location and prevalence of effective humoral responses, we used a serologic analysis based on serum neutralization of a panel of diverse HIV-1 isolates<sup>20</sup>. Sera from a cohort that had been infected for 2–3 years and from another that had been infected for more than 5 years were assessed on a panel of 21 diverse HIV-1 isolates, and the neutralization phenotypes assigned to 12 prototypic antibody-neutralization fingerprints (Fig. 6c, Extended Data Fig. 8a, b). We then mapped the responses to the surface of the mature closed HIV-1 Env spike (Extended Data Fig. 8c, d). The most prevalent response corresponded to the glycan-V3 epitope epitomized by antibody PGT128. CD4-binding site-directed responses, 8ANC195 responses, V1V2-directed responses, and 35O22 responses were also prevalent after 5+ years. Overall, responses in both cohorts were generally in good agreement with each other, indicating little evolution in the location or prevalence of effective neutralizing responses between 2–3 and 5+ years. Notably, when mapping Env sites of vulnerability, the majority of prevalent sites corresponded to Env surfaces covered by N-linked glycosylation and/or of high sequence variability. Indeed, both PGT122 and 35O22 co-crystallized here recognize N-linked glycan, and they both utilize framework 3 insertions, in the light chain for PGT122 and in the heavy chain for 35O22 (Extended Data Fig. 9).

### Viral evasion and immune recognition

In addition to merging virus and host-cell membranes, viral fusion machines must contend with antibody-mediated neutralization. With RSV, peak infection occurs at 5–10 months of life, as maternal antibodies wane; with influenza virus, natural infection elicits strain-specific antibodies, and evasion occurs seasonally on a global scale. HIV-1, however, confronts the immune system in each individual directly, often presenting high titre of Env antigens over years of chronic infection. These differences in evasion are reflected by structural differences in the fusion machines. The structure of the HIV-1 Env spike revealed here allows the molecular trickery behind single-spike entry<sup>43</sup>, glycan shielding<sup>3</sup> and conformational masking<sup>44</sup> to be visualized at the atomic level (Extended Data Fig. 10). Thus, avoidance of antibody avidity<sup>45</sup> through the ability of a single HIV-1 spike to fuse viral and target cell membranes<sup>43</sup> is likely to be assisted by membrane proximity of the co-receptor and membrane association of the membrane-proximal external region (MPER; Fig. 3); despite these differences, the HIV-1 Env spike appears to share mechanism and topology with other type I fusion machines (Fig. 4). In terms of glycan shielding<sup>3</sup>, we have modelled the structure of a fully assembled glycan shield for BG505, a tier II-transmitted founder virus<sup>46</sup> (Fig. 5). Although glycan masking appears to be complete at the HIV-1 spike apex, closer to the viral membrane ‘holes’ in the glycan shield are observed. And with conformational masking<sup>44</sup>, evasion is optimal for the pre-fusion mature closed state, with receptor-binding unmasking conserved glycan-free surfaces (Extended Data Fig. 7). Despite extraordinary glycosylation and sequence variation, the human immune system seems to be up to the challenge of generating HIV-1-neutralizing antibodies (Fig. 6). We note that recognition of glycosylation appears to be a trait common only to HIV-1-neutralizing antibodies and that

both broadly neutralizing HIV-1 and influenza virus antibodies tolerate epitope sequence variation (Fig. 6b). The structure of the HIV-1 Env spike described here thus reveals not only commonalities in entry and evasion with other type I fusion machines, but also commonalities in recognition by the human immune system. It remains to be seen whether an effective vaccine against HIV-1 can be developed by using the atomic-level detail provided here, which should allow for immunogen-design strategies such as conformational stabilization<sup>47</sup> and nanoparticle delivery<sup>48</sup>; additionally, antibody-type and ontogeny-specific strategies may be required, and template ontogenies are becoming available for some of the more commonly elicited HIV-1-neutralizing antibodies (Extended Data Fig. 8d), such as those against the CD4-binding site<sup>49</sup> and V1V2 site<sup>50</sup>.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 4 June; accepted 1 September 2014.

Published online 8 October 2014.

1. The Joint United Nations Programme on HIV/AIDS. UNAIDS report on the global AIDS epidemic 2013 [http://www.unaids.org/en/media/unaids/contentassets/documents/epidemiology/2013/gr2013/unaids\\_global\\_report\\_2013\\_en.pdf](http://www.unaids.org/en/media/unaids/contentassets/documents/epidemiology/2013/gr2013/unaids_global_report_2013_en.pdf) (2013).
2. Wyatt, R. & Sodroski, J. The HIV-1 envelope glycoproteins: fusogens, antigens, and immunogens. *Science* **280**, 1884–1888 (1998).
3. Wei, X. *et al.* Antibody neutralization and escape by HIV-1. *Nature* **422**, 307–312 (2003).
4. Leonard, C. K. *et al.* Assignment of intrachain disulfide bonds and characterization of potential glycosylation sites of the type 1 recombinant human immunodeficiency virus envelope glycoprotein (gp120) expressed in Chinese hamster ovary cells. *J. Biol. Chem.* **265**, 10373–10382 (1990).
5. Cimbardo, R. *et al.* Tyrosine sulfation in the second variable loop (V2) of HIV-1 gp120 stabilizes V2–V3 interaction and modulates neutralization sensitivity. *Proc. Natl Acad. Sci. USA* **111**, 3152–3157 (2014).
6. Li, Y. *et al.* Effects of inefficient cleavage of the signal sequence of HIV-1 gp 120 on its association with calnexin, folding, and intracellular transport. *Proc. Natl Acad. Sci. USA* **93**, 9606–9611 (1996).
7. Kwong, P. D. *et al.* Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* **393**, 648–659 (1998).
8. Chan, D. C., Fass, D., Berger, J. M. & Kim, P. S. Core structure of gp41 from the HIV envelope glycoprotein. *Cell* **89**, 263–273 (1997).
9. Weissenhorn, W., Dessen, A., Harrison, S. C., Skehel, J. J. & Wiley, D. C. Atomic structure of the ectodomain from HIV-1 gp41. *Nature* **387**, 426–430 (1997).
10. Julien, J. P. *et al.* Crystal structure of a soluble cleaved HIV-1 envelope trimer. *Science* **342**, 1477–1483 (2013).
11. Lyumkis, D. *et al.* Cryo-EM structure of a fully glycosylated soluble cleaved HIV-1 envelope trimer. *Science* **342**, 1484–1490 (2013).
12. Walker, L. M. *et al.* Broad neutralization coverage of HIV by multiple highly potent antibodies. *Nature* **477**, 466–470 (2011).
13. Huang, J. *et al.* Broad and potent neutralization of HIV-1 by a human antibody that recognizes an intersubunit site on the envelope glycoprotein. *Nature* <http://dx.doi.org/10.1038/nature13601> (3 September 2014).
14. Sanders, R. W. *et al.* Stabilization of the soluble, cleaved, trimeric form of the envelope glycoprotein complex of human immunodeficiency virus type 1. *J. Virol.* **76**, 8875–8889 (2002).
15. Julien, J. P. *et al.* Asymmetric recognition of the HIV-1 trimer by broadly neutralizing antibody PG9. *Proc. Natl Acad. Sci. USA* **110**, 4351–4356 (2013).
16. Sanders, R. W. *et al.* A next-generation cleaved, soluble HIV-1 Env trimer, BG505 SOSIP.664 gp140, expresses multiple epitopes for broadly neutralizing but not non-neutralizing antibodies. *PLoS Pathog.* **9**, e1003618 (2013).
17. Ringe, R. P. *et al.* Cleavage strongly influences whether soluble HIV-1 envelope glycoprotein trimers adopt a native-like conformation. *Proc. Natl Acad. Sci. USA* **110**, 18256–18261 (2013).
18. Munro, J. B. *et al.* Conformational dynamics of single HIV-1 envelope trimers on the surface of native virions. *Science* <http://dx.doi.org/10.1126/science.1254426> (2014).
19. Julien, J. P. *et al.* Broadly neutralizing antibody PGT121 allosterically modulates CD4 binding via recognition of the HIV-1 gp120 V3 base and multiple surrounding glycans. *PLoS Pathog.* **9**, e1003342 (2013).
20. Georgiev, I. S. *et al.* Delineating antibody recognition in polyclonal sera from patterns of HIV-1 isolate neutralization. *Science* **340**, 751–756 (2013).
21. Mao, Y. *et al.* Subunit organization of the membrane-bound HIV-1 envelope glycoprotein trimer. *Nature Struct. Mol. Biol.* **19**, 893–899 (2012).
22. Pancera, M. *et al.* Structure of HIV-1 gp120 with gp41-interactive region reveals layered envelope architecture and basis of conformational mobility. *Proc. Natl Acad. Sci. USA* **107**, 1166–1171 (2010).
23. Finzi, A. *et al.* Topological layers in the HIV-1 gp120 inner domain regulate gp41 interaction and CD4-triggered conformational transitions. *Mol. Cell* **37**, 656–667 (2010).

24. Buzon, V. *et al.* Crystal structure of HIV-1 gp41 including both fusion peptide and membrane proximal external regions. *PLoS Pathog.* **6**, e1000880 (2010).
25. Caffrey, M. *et al.* Three-dimensional solution structure of the 44 kDa ectodomain of SIV gp41. *EMBO J.* **17**, 4572–4584 (1998).
26. Nishikawa, K., Ooi, T., Saito, N. & Isogai, Y. Tertiary structure of proteins. 1. Representation and computation of conformations. *J. Phys. Soc. Jpn.* **32**, 1331–1337 (1972).
27. Bartsaghi, A., Merk, A., Borgnia, M. J., Milne, J. L. & Subramaniam, S. Prefusion structure of trimeric HIV-1 envelope glycoprotein determined by cryo-electron microscopy. *Nature Struct. Mol. Biol.* **20**, 1352–1357 (2013).
28. Liu, J., Bartsaghi, A., Borgnia, M. J., Sapiro, G. & Subramaniam, S. Molecular architecture of native HIV-1 gp120 trimers. *Nature* **455**, 109–113 (2008).
29. White, T. A. *et al.* Molecular architectures of trimeric SIV and HIV-1 envelope glycoproteins on intact viruses: strain-dependent variation in quaternary structure. *PLoS Pathog.* **6**, e1001249 (2010).
30. Huang, C. C. *et al.* Structure of a V3-containing HIV-1 gp120 core. *Science* **310**, 1025–1028 (2005).
31. McLellan, J. S. *et al.* Structure of RSV fusion glycoprotein trimer bound to a prefusion-specific neutralizing antibody. *Science* **340**, 1113–1117 (2013).
32. Yuan, W., Craig, S., Si, Z., Farzan, M. & Sodroski, J. CD4-induced T-20 binding to human immunodeficiency virus type 1 gp120 blocks interaction with the CXCR4 coreceptor. *J. Virol.* **78**, 5448–5457 (2004).
33. Huang, C. C. *et al.* Structures of the CCR5 N terminus and of a tyrosine-sulfated antibody with HIV-1 gp120 and CD4. *Science* **317**, 1930–1934 (2007).
34. Moore, J. P., McKeating, J. A., Weiss, R. A. & Sattentau, Q. J. Dissociation of gp120 from HIV-1 virions induced by soluble CD4. *Science* **250**, 1139–1142 (1990).
35. Wilson, I. A., Skehel, J. J. & Wiley, D. C. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. *Nature* **289**, 366–373 (1981).
36. Bullough, P. A., Hughson, F. M., Skehel, J. J. & Wiley, D. C. Structure of influenza haemagglutinin at the pH of membrane fusion. *Nature* **371**, 37–43 (1994).
37. McLellan, J. S., Yang, Y., Graham, B. S. & Kwong, P. D. Structure of respiratory syncytial virus fusion glycoprotein in the postfusion conformation reveals preservation of neutralizing epitopes. *J. Virol.* **85**, 7788–7796 (2011).
38. Weissenhorn, W., Carfi, A., Lee, K. H., Skehel, J. J. & Wiley, D. C. Crystal structure of the Ebola virus membrane fusion subunit, GP2, from the envelope glycoprotein ectodomain. *Mol. Cell* **2**, 605–616 (1998).
39. Lee, J. E. *et al.* Structure of the Ebola virus glycoprotein bound to an antibody from a human survivor. *Nature* **454**, 177–182 (2008).
40. Colman, P. M. & Lawrence, M. C. The structural biology of type I viral membrane fusion. *Nature Rev. Mol. Cell Biol.* **4**, 309–319 (2003).
41. Wyatt, R. *et al.* The antigenic structure of the HIV gp120 envelope glycoprotein. *Nature* **393**, 705–711 (1998).
42. Hraber, P. *et al.* Prevalence of broadly neutralizing antibody responses during chronic HIV-1 infection. *AIDS* **28**, 163–169 (2014).
43. Yang, X., Kurteva, S., Ren, X., Lee, S. & Sodroski, J. Stoichiometry of envelope glycoprotein trimers in the entry of human immunodeficiency virus type 1. *J. Virol.* **79**, 12132–12147 (2005).
44. Kwong, P. D. *et al.* HIV-1 evades antibody-mediated neutralization through conformational masking of receptor-binding sites. *Nature* **420**, 678–682 (2002).
45. Klein, J. S. & Bjorkman, P. J. Few and far between: how HIV may be evading antibody avidity. *PLoS Pathog.* **6**, e1000908 (2010).
46. Wu, X. *et al.* Neutralization escape variants of human immunodeficiency virus type 1 are transmitted from mother to infant. *J. Virol.* **80**, 835–844 (2006).
47. McLellan, J. S. *et al.* Structure-based design of a fusion glycoprotein vaccine for respiratory syncytial virus. *Science* **342**, 592–598 (2013).
48. Kanekiyo, M. *et al.* Self-assembling influenza nanoparticle vaccines elicit broadly neutralizing H1N1 antibodies. *Nature* **499**, 102–106 (2013).
49. Liao, H.-X. *et al.* Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus. *Nature* **496**, 469–476 (2013).
50. Doria-Rose, N. A. *et al.* Developmental pathway for potent V1V2-directed HIV-neutralizing antibodies. *Nature* **509**, 55–62 (2014).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank J. Binley for the JR-FL plasmid used in smFRET; D. Burton and W. Koff for PGT122; J. Chrzas and SER-CAT staff for assistance with data collection; B. Graham for discussions on RSV; R. Sanders for furin plasmid; R. Schwartz and the Vaccine Production Program for antibody VRC01; J. Sodroski for information on C34-Ig; I. Wilson for discussions on pre-fusion Env structure; Z. Yang for assistance with cloning; G. Nabel, G. Scott and members of the Structural Biology Section and Structural Bioinformatics Core, Vaccine Research Center, for discussions and comments on the manuscript; and the WCMC/AMC/TSRI HIVRAD team for their contributions to the design and validation of near-native mimicry for soluble BG505 SOSIP.664 trimers. We thank J. Baalwa, D. Ellenberger, F. Gao, B. Hahn, K. Hong, J. Kim, F. McCutchan, D. Montefiori, L. Morris, J. Overbaugh, E. Sanders-Buell, G. Shaw, R. Swanstrom, M. Thomson, S. Tovanabutra, C. Williamson, and L. Zhang for contributing the HIV-1 envelope plasmids used in our neutralization panel. The authors acknowledge the contributions of the Center for HIV/AIDS Vaccine Discovery (CHAVI) Clinical Core Team for recruiting study participants and carrying out all aspects of the CHAVI001 and CHAVI008 protocols at Chapel Hill, North Carolina (J. Eron); Blantyre, Malawi (J. Kumwenda, T. Taha); Lilongwe, Malawi (I. Hoffman, G. Kamanga); Johannesburg, South Africa (H. Rees); Durban, South Africa (S. Abdool Karim); Moshi, Tanzania (S. Noel, S. Kapiga, J. Crump); and London, UK (S. Fidler). Support for this work was provided by the Intramural Research Program of the Vaccine Research Center, National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health (NIH); and by grants from the Division of AIDS, NIAID, NIH (R21-AI100696, CHAVI-AI0678501, and CHAVI-Immunogen Discovery-AI100645), from the National Institutes of General Medical Sciences (P01-GM56550 and R01-GM098859), from the Irvington Fellows Program of the Cancer Research Program, and from the International AIDS Vaccine Initiative's (IAVI's) Neutralizing Antibody Consortium. IAVI's work is made possible by support from many donors including: the Bill & Melinda Gates Foundation; the Ministry of Foreign Affairs of Denmark; Irish Aid; the Ministry of Finance of Japan; the Ministry of Foreign Affairs of the Netherlands; the Norwegian Agency for Development Cooperation (NORAD); the UK Department for International Development (DFID); and the United States Agency for International Development (USAID). The full list of IAVI donors is available at <http://www.iavi.org>. Use of sector 22 (Southeast Region Collaborative Access Team) at the Advanced Photon Source was supported by the US Department of Energy, Basic Energy Sciences, Office of Science, under contract number W-31-109-Eng-38.

**Author Contributions** M.P., T.Z., A.D. and P.D.K. determined the trimer structure, with M.P. heading structure determination; and with T.Z. assisting with solution and refinement, A.D. with protein production, and P.D.K. with data collection. I.S.G. and C.S. performed bioinformatics analysis, J.G. performed conformational analysis, P.A., G.-Y.C., G.O., G.B.E.S.-J. and J.S. performed antigenic and mechanistic analyses, R.T.B., M.G.J., M.K.L., N.T., Y.Y., B.Z., M.S.C., B.F.H., J.R.M. and L.M. performed cohort analysis, J.B.M., S.C.B. and W.M. performed smFRET analysis, and J.H. and M.C. provided antibody 35022. M.P. and P.D.K. wrote the paper, on which all principal investigators commented.

**Author Information** Coordinates and structure factors for BG505 SOSIP.664 in complex with PGT122 and 35022 Fabs have been deposited with the Protein Data Bank under accession code 4TVP. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.D.K. ([pdkwong@nih.gov](mailto:pdkwong@nih.gov)).

## METHODS

**BG505 SOSIP.664 expression and purification.** The crystallized HIV-1 Env construct from strain BG505 was generated following published reports<sup>10,15,16</sup>, using BG505 GenBank accession numbers ABA61516 and DQ208458<sup>16</sup>; including the 'SOS' mutations (A501C, T605C), the isoleucine to proline mutation at residue 559 (I559P), and the glycan site at residue 332 (T332N); mutating the cleavage site to 6R (REKR to RRRRRR); and truncating the C terminus to residue 664 (all HIV-1 Env numbering according to the HX nomenclature). This construct is referred to as BG505 SOSIP.664 throughout this entire manuscript.

The BG505 SOSIP.664 construct was co-transfected with furin in HEK 293 GnTI<sup>-/-</sup> cells using 600 µg of BG505 SOSIP.664 and 150 µg of furin plasmid DNAs as described previously<sup>16</sup>. Transfection supernatants were harvested after 7 days, and passed over either a 2G12 antibody- or VRC01 antibody-affinity column. After washing with phosphate-buffered saline (PBS), bound proteins were eluted with 3 M MgCl<sub>2</sub>, 10 mM Tris pH 8.0. The eluate was concentrated to less than 4 ml with Centricon-70 and applied to a Superdex 200 column, equilibrated in 5 mM HEPES, pH 7.5, 150 mM NaCl, 0.02% azide. The peak corresponding to trimeric HIV-1 Env was identified, pooled, concentrated and used immediately or flash-frozen in liquid nitrogen and stored at -80 °C.

**Fab expression and purification.** PGT122 and 35O22 IgGs were expressed as previously described<sup>51</sup>. Heavy chain plasmids containing an HRV3C cleavage site in the hinge region were co-transfected with light chain plasmids in 293F (35O22) or GnTI<sup>-/-</sup> (PGT122, which is glycosylated) using TrueFect-Max transfection reagent (United Biosystems) according to manufacturer's protocol. Cultures were fed with fresh 293FreeStyle media (Life Technologies) 4 h post-transfection and with HyClone SFM4HEK293 enriched medium (HyClone) containing valproic acid (4 mM final concentration) 24 h after transfection. Cultures were then incubated at 33 °C for 6 days, and supernatants harvested and passed over a protein A affinity column. After PBS wash and low pH elution, pH of eluate was neutralized with 1 M Tris pH 8.5. Fabs were obtained using HRV3C digestion and collecting flow-through from protein A column to remove Fc fraction. Fabs were further purified over Superdex 200 in 5 mM HEPES, pH 7.5, 150 mM NaCl, 0.02% azide.

**Ternary complex preparation.** PGT122 and 35O22 Fabs were added to a solution of purified trimeric BG505 SOSIP.664 in fivefold molar excess for 30 min at room temperature. The complex was then partially deglycosylated by adding Endo H (50 µl) for 1 h at room temperature in the gel filtration buffer. The complex was then purified over gel filtration equilibrated in 5 mM HEPES, pH 7.5, 150 mM NaCl, 0.02% azide. Fractions were pooled, concentrated down to 5–10 OD<sub>280 nm</sub> per ml and used immediately for crystal screening or flash frozen in liquid nitrogen and kept at -80 °C until further use.

**Crystallization screening.** The ternary complex was screened for crystallization using 576 conditions from Hampton, Wizard and Precipitant Synergy<sup>52</sup> screens using a Cartesian Honeybee crystallization robot as described previously<sup>51</sup> and a mosquito robot using 0.1 µl of reservoir solution and 0.1 µl of protein solution. Crystals suitable for structural determination were identified robotically in 0.2 M Li<sub>2</sub>SO<sub>4</sub>, 6.65% PEG 1500, 20% isopropanol and 0.1 M sodium acetate pH 5.5. Crystals were reproduced in hanging droplets containing 0.5 µl of reservoir solution and 0.5 µl of protein solution. Optimal crystallization conditions were obtained in 16% isopropanol, 5.32% PEG 1500, 0.2 M Li<sub>2</sub>SO<sub>4</sub>, 0.1 M Na acetate pH 5.5. Crystals were cryoprotected in a solution of 15% 2R3R-butanediol, 5% isopropanol, 0.2 M Li<sub>2</sub>SO<sub>4</sub>, 6.65% PEG 1500, 0.1 M sodium acetate pH 5.5, and flash-frozen after covering with paratone N<sup>53</sup>. Data were collected at a wavelength of 1.00 Å at the SER-CAT beamline ID-22 (Advanced Photon Source, Argonne National Laboratory).

**X-ray data collection, structure solution and model building.** Diffraction data were processed with the HKL2000 suite<sup>54</sup>. The data were corrected for anisotropy using the anisotropy server <http://services.mbi.ucla.edu/anisotropy/> with truncations to 3.5 Å, 3.5 Å, 3.1 Å along *a*, *b* and *c* axes, respectively. Structure solution was obtained with Phaser using gp120 (PDB ID 4J6R<sup>20</sup>), PGT122 (PDB ID 4JY5<sup>19</sup>) and 35O22 Fv (PDB ID 4TOY<sup>13</sup>) as search models. Refinement was carried out with Phenix<sup>55</sup> imposing PGT122, 35O22 and gp120 model-based refinement restraint during initial round of refinement. Model building was carried out with Coot<sup>56</sup>. The Ramachandran plot as determined by MOLPROBITY<sup>57</sup> showed 92.66% of all residues in favoured regions and 99.03% of all residues in allowed regions. Data collection and refinement statistics are shown in Extended Data Table 1.

**Preparation of fluorescently labelled virus.** Fluorescently labeled virus was prepared as described<sup>18</sup>. Briefly, for site-specific incorporation of fluorophores the Q3 (GQQQLG) and A1 (GDSLDMLEWSLM) peptides were inserted into the V1 and V4 loops of HIV-1 JR-FL gp160 at positions 136 and 404 (HXB2 numbering), respectively. Virus for smFRET imaging was generated by cotransfecting HEK293 cells with a 40:1 ratio of wild-type HIV-1 JR-FL gp160 plasmid pCAGGS to HIV-1 JR-FL gp160 plasmid containing the Q3 and A1 labelling peptides, in addition to pNL4-3 Δenv ART. The virus was harvested 24 h post-transfection, concentrated by centrifugation, and fluorescently labelled with donor and acceptor fluorophores

through incubation with 0.5 µM Cy3B-cadaverine, 0.5 µM Cy5(4S)COT-CoA, 0.65 µM transglutaminase<sup>58</sup> (Sigma), and 5 µM AcpS<sup>59</sup> overnight at room temperature. The AcpS enzyme and the CoA-conjugated fluorophore were prepared as described<sup>59</sup>. DSPE-PEG<sub>2,000</sub>-biotin (Avanti) was then added to the reaction at a final concentration of 6 µM (0.02 mg ml<sup>-1</sup>), and the labelled virus was purified by ultracentrifugation for 1 h at 150,000g over a 6–18% Optiprep (Sigma) gradient.

**smFRET data acquisition and analysis.** smFRET data were acquired and analysed as described<sup>18</sup>. Fluorescently labelled virions were immobilized on streptavidin-coated quartz microscope slides and imaged on a prism-based total internal reflection fluorescence microscope. The donor fluorophore was excited by a 532-nm laser (Laser Quantum). The donor and acceptor fluorescence emissions were collected through a ×60 water objective (Nikon), split by a 650DCXR dichroic filter (Chroma), and focused on parallel EMCCD cameras (Photometrics). Movies were recorded at 25 frames per s for 40 s. smFRET imaging was performed in buffer containing 50 mM Tris pH 7.5, 100 mM NaCl, a cocktail of triplet-state quenchers<sup>60</sup>, and 2 mM protocatechuic acid and 8 nM protocatechuate 3,4-deoxygenase to remove molecular oxygen<sup>61</sup>. Where indicated, surface-bound viruses were incubated with 0.1 mg ml<sup>-1</sup> PGT122 and/or 0.1 mg ml<sup>-1</sup> 35O22 antibody.

All data analysis was performed using custom written Matlab software. Fluorescence trajectories were extracted from the movies, and used to calculate FRET efficiency according to  $FRET = IA/(ID + IA)$ , where *IA* and *ID* represent fluorescence intensities of the acceptor and donor fluorophores, respectively. smFRET trajectories were identified for analysis on basis of their displaying sufficient signal-to-noise and fluorophore lifetime. FRET trajectories were compiled into histograms, which were fit to the sum of three Gaussian distributions in Matlab. smFRET revealed that the HIV-1 Env is conformationally dynamic, transitioning between three distinct conformational states. Response to various ligands identified the low-FRET state as the closed unliganded conformation of HIV-1 Env and the intermediate- and high-FRET states as the activated conformations stabilized by coreceptor and CD4 binding.

**Binding studies using biolayer interferometry.** A forteBio Octet Red384 instrument was used to measure binding of BG505 SOSIP.664 and BG505 gp120 molecules to a panel of antibodies (VRC01, VRC03, b6, b12, F105, PGT122, PGT128, PGT135, 2G12, 8ANC195, 17b, 2.2C, 412d, 48D, 447-52D, PG9, PG16, PGT145, VRC26.09, 35O22, PGT151) and CD4 Ig. All the assays were performed with agitation set to 1,000 r.p.m. in PBS buffer supplemented with 1% bovine serum albumin (BSA) in order to minimize nonspecific interactions. The final volume for all the solutions was 40–50 µl per well. Assays were performed at 30 °C in solid black tilted-bottom 384-well plates (Geiger Bio-One). Human antibodies (40–50 µg ml<sup>-1</sup>) in PBS buffer was used to load anti-human IgG Fc capture (AHC) probes for 600 s. Typical capture levels were between 1 and 1.5 nm, and variability within a row of eight tips did not exceed 0.1 nm. Biosensor tips were then equilibrated for 180 s in PBS/1% BSA buffer before binding assessment of the BG505 SOSIP.664 and BG505 gp120 molecules in solution for 300 s; binding was then allowed to dissociate for 300 s. Parallel correction to subtract systematic baseline drift was carried out by subtracting the measurements recorded for a sensor without monoclonal antibody incubated in PBS/1% BSA. Data analyses were carried out using Octet software, version 8.0.

**Difference distance analysis.** Difference distance matrices<sup>26</sup> were produced by distance sorting atom positions and plotting with the program DDMP<sup>62</sup>.

**Surface plasmon resonance analysis.** Affinities and kinetics of binding of antibodies 35O22 and PGT151 to BG505 SOSIP.664 soluble trimer were assessed by surface plasmon resonance on a Biacore T-200 (GE Healthcare) at 20 °C with buffer HBS-EP+ (10 mM HEPES, pH 7.4, 150 mM NaCl, 3 mM EDTA, and 0.05% surfactant P-20). In general, mouse anti-human Fc antibody was first immobilized onto two flow cells on a CM5 chip at ~10,000 response units (RU) with standard amine coupling protocol (GE Healthcare). Either CD4-Ig, 2G12 IgG or 17b IgG was then captured on both flow cells by flowing over a 200 nM solution at 5 µl min<sup>-1</sup> flow rate for two minutes. This was followed by a 1-min injection of 1 µM human Fc on both flow cells to block unliganded mouse anti-human Fc antibody. The captured 2G12, CD4 or 17b were used to immobilize BG505 SOSIP.664 trimer on only one flow cell, with no trimer captured on the other flow cell (reference cell). For capturing with 2G12 or CD4-Ig, 500 nM of unliganded trimer was used, whereas, a complex of 500 nM trimer + 1,500 nM sCD4 was used for capturing with 17b. Antibody Fab fragment solutions in HBS-EP+ buffer, at twofold dilutions starting from 885 nM, 600 nM and 460 nM for 35O22, PGT151 and PGT145, respectively, were injected over the captured trimer channel and the reference channel at a flow rate of 50 µl min<sup>-1</sup> for 2 min and allowed to dissociate for 3–30 min depending on the rate of dissociation of each interaction. The cells were regenerated with two 10 µl injections of 3.0 M MgCl<sub>2</sub> at a flow rate of 100 µl min<sup>-1</sup>. Blank sensorgrams were obtained by injection of same volume of HBS-EP+ buffer without antibody Fab fragments. Sensorgrams of the concentration series were corrected with corresponding blank curves and fitted globally with Biacore T200 evaluation software using a 1:1 Langmuir model of binding. The stoichiometry of binding of antibodies to the trimer



were estimated by normalizing the maximal response ( $R_{\max}$ ) values to the amount of trimer captured and performing linear regression analysis using the  $R_{\max}$  values for the antibodies with known stoichiometries.

**Modelling of missing loops, side chains and the N-linked glycan shield.** Missing loops not defined in the HIV-1 Env trimer crystal structure (V2 and V4) were modelled using Loopy<sup>63</sup>. Missing side chains were modelled with Scap<sup>64</sup>.

To model the N-linked glycan shield, we first determined all possible N-linked sequons in the BG505 HIV-1 Env trimer sequence (28 sequons). All glycans observed in the structure were removed before modelling. A conserved glycan 241<sub>gp120</sub>, not present in the BG505 sequence, was added. A single asparagine residue in each sequon was targeted for computational N-linked glycan addition using a series of oligomannose 9 rotamer libraries at different resolutions. In constructing the rotamer libraries, the asparagine side chain rotamers were also considered. To avoid a combinatorial explosion in the search space, select torsion angles in the oligomannose 9 rotamer libraries were allowed to vary in increments between 30–60 degrees. We used an overlap factor (*ofac*) to screen for clashes between the sugar moieties and the trimer structure. The *ofac* between two non-bonded atoms is defined as the distance between two atoms divided by the sum of their van der Waals radii. For the modelling carried out here, we set the *ofac* to a value of 0.60. For sterically occluded positions, the *ofac* was set to 0.55. To remove steric bumps between sugar moieties, all models were subjected to 100 cycles of conjugate gradient energy minimization using the GLYCAM<sup>65</sup> force field in Amber12<sup>66</sup> with a distance-dependent dielectric.

**Mapping sequence variability onto trimer structure.** For each of HIV-1 Env, influenza virus HA, and RSV F, residue sequence variability was computed as the Shannon entropy for each residue position, based on representative sets of 3,943 HIV-1 strains, 4,467 influenza virus strains, and 212 RSV strains, respectively. Residues were coloured based on the computed entropy values, on a scale of white (conserved) to purple (variable).

**Chronically infected cohort information.** In the CHAVI 001 cohort, high-risk subjects were screened for HIV-1 infection by ELISA, western blotting, and plasma RNA to recruit individuals with acute HIV infection, who were then followed for ~2 years until plasma neutralization breadth developed<sup>67</sup>. In addition, a group of individuals were enrolled in the CHAVI 001 or CHAVI 008 cohorts who were chronically infected with HIV-1 strains clade A, B or C, and were screened for plasma neutralization breadth. The trial participants were enrolled at sites in Tanzania, South Africa, Malawi, the United States, and the United Kingdom<sup>68</sup>. Both CHAVI001 and CHAVI008 protocols were approved by the institutional review boards of each of the participating institutions where blood samples were received or processed for analysis, and informed consent was obtained from all subjects.

**Serum neutralization fingerprinting analysis.** The prevalence of effective neutralizing responses against HIV-1 Env in cohorts from 2–3 and 5+ years post-infection was estimated using a neutralization fingerprinting approach, as described previously<sup>20</sup>. Briefly, serum neutralization over a set of 21 diverse viral strains was compared to neutralization of the same viruses by a set of broadly neutralizing antibodies grouped into 12 epitope-specific antibody clusters. For each serum, the relative prevalence of each of the 12 antibody specificities was estimated by representing serum neutralization as a linear combination of the monoclonal specificities, with prevalence values of 0.2 deemed as positive. Sera with less than 30% breadth on the 21-virus panel as well as sera with high residual values from the computation (data not shown) were not included in the analysis. For mapping prevalence values onto the BG505 SOSIP.664 structure, residues part of multiple antibody epitopes were coloured according to the respective antibody specificity with the highest prevalence in the 5+ years cohort. Antibody neutralization was measured using single-round-of-infection HIV-1 Env-pseudoviruses and TZM-bl target cells, as described previously<sup>69</sup>. Neutralization curves were fit by nonlinear regression using a 5-parameter hill slope equation as previously described<sup>69</sup>.

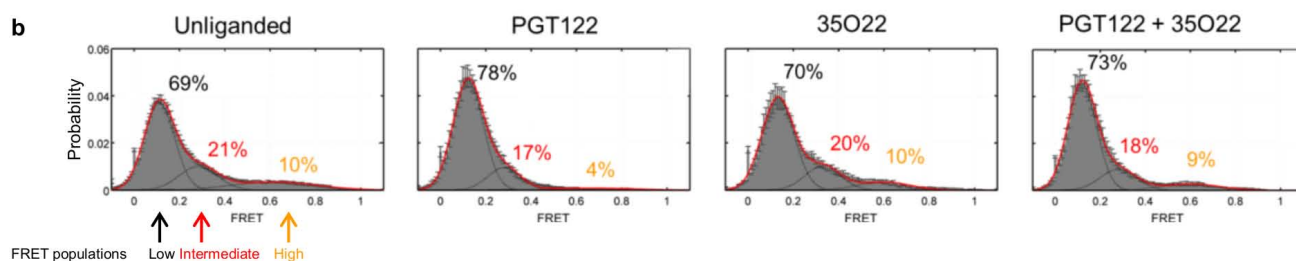
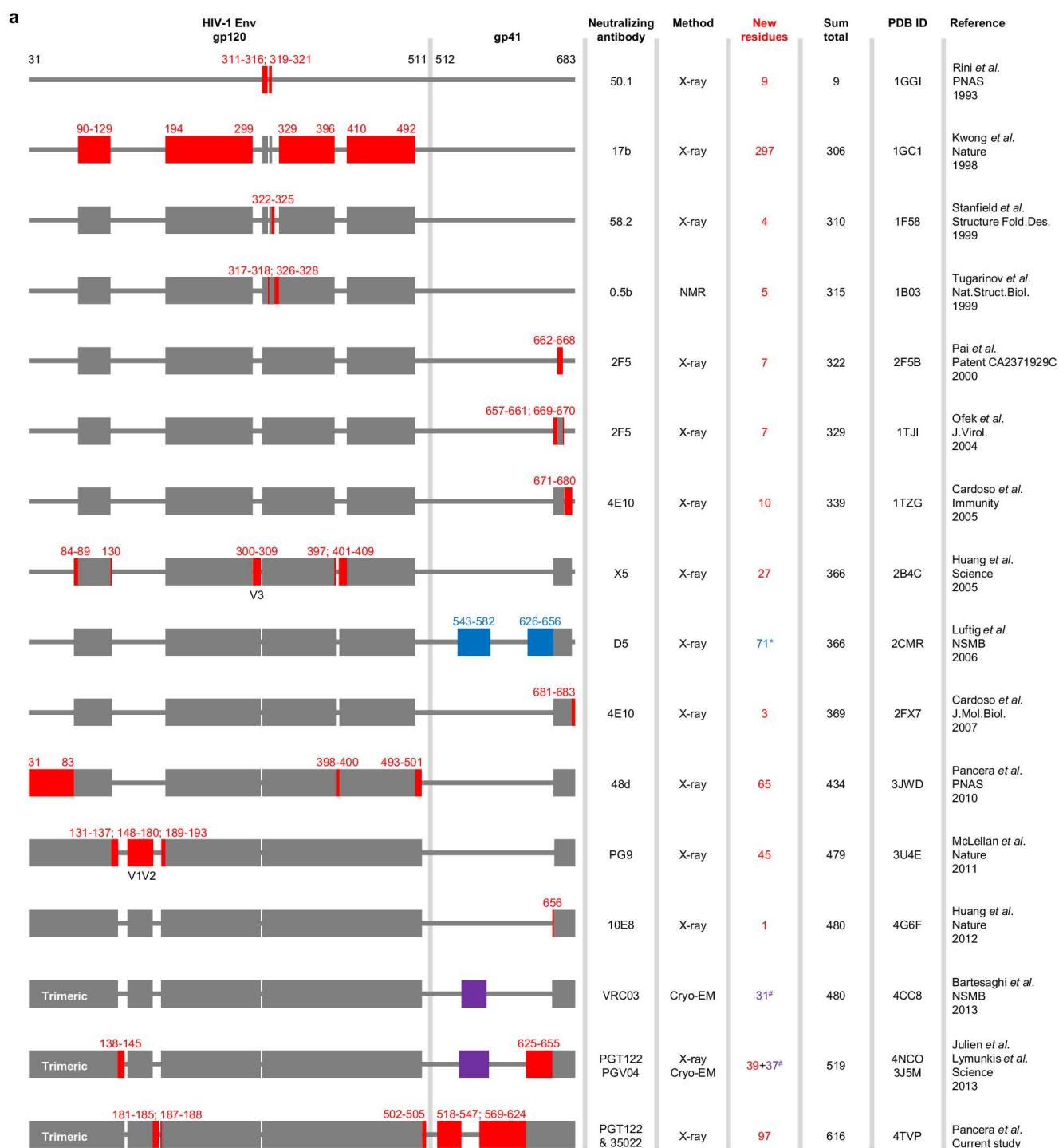
**Epitope analysis for HIV-1 Env, influenza virus HA and RSV F antibodies.** Glycan usage and average residue entropy were calculated for seven representative HIV-1 Env (VRC01, b12, CD4, 8ANC195, PG9, PGT122, 2G12 and 35O22)<sup>7,51,70–73</sup>, four representative influenza virus HA (2D1, C05, F10 and CR8043)<sup>74–77</sup>, and three representative RSV F (D25, Motavizumab and 101F)<sup>31,78,79</sup> epitopes based on their respective crystal structures. The selection of the flu antibodies was done as follows: F10 (stem targeting) and C05 (head targeting) were selected based on their cross-neutralizing ability for group 1 and group 2 of influenza virus A. CR8043 (group 2 specific) and 2D1 (H1 specific), which target distinct regions from F10 and C05 at the stem and head of the HA respectively, were also selected for epitope analysis. An antigen residue was defined as an epitope residue if it had a non-zero buried surface area in the crystal structure. The fraction of glycan surface area in an epitope was calculated as the buried surface area of epitope glycans divided by the buried surface area of the full epitope. Unpaired nonparametric Mann–Whitney test<sup>80</sup> was used to quantify the statistical difference between glycan fraction or average residue entropy for HIV-1 versus influenza virus or RSV antibody epitopes.

**Figures.** Structure figures were prepared using PYMOL<sup>81</sup>. PDB IDs are referenced throughout except for 2HMG<sup>82</sup>, 1QU1<sup>83</sup>, 4MMU<sup>47</sup>, 3RRT<sup>37</sup>, 3CSY<sup>39</sup> and 2EBO<sup>84</sup> in Fig. 4; 2YP7<sup>85</sup> and 4JHW<sup>31</sup> in Figs 5 and 6 and PDB ID 3H1I<sup>86</sup> in Extended Data Table 2.

**Interfaces.** Interactive surfaces were obtained from PISA (<http://www.ebi.ac.uk/pdbe/pisa/>).

51. McLellan, J. S. *et al.* Structure of HIV-1 gp120 V1/V2 domain with broadly neutralizing antibody PG9. *Nature* **480**, 336–343 (2011).
52. Majeed, S. *et al.* Enhancing protein crystallization through precipitant synergy. *Structure* **11**, 1061–1070 (2003).
53. Kwong, P. D. & Liu, Y. Use of cryoprotectants in combination with immiscible oils for flash cooling macromolecular crystals. *J. Appl. Cryst.* **32**, 102–105 (1999).
54. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
55. Adams, P. D. *et al.* Recent developments in the PHENIX software for automated crystallographic structure determination. *J. Synchrotron Radiat.* **11**, 53–55 (2004).
56. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
57. Davis, I. W., Murray, L. W., Richardson, J. S. & Richardson, D. C. MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res.* **32**, W615–W619 (2004).
58. Lin, C. W. & Ting, A. Y. Transglutaminase-catalyzed site-specific conjugation of small-molecule probes to proteins *in vitro* and on the surface of living cells. *J. Am. Chem. Soc.* **128**, 4542–4543 (2006).
59. Zhou, Z. *et al.* Genetically encoded short peptide tags for orthogonal protein labeling by Sfp and AcpS phosphopantetheinyl transferases. *ACS Chem. Biol.* **2**, 337–346 (2007).
60. Dave, R., Terry, D. S., Munro, J. B. & Blanchard, S. C. Mitigating unwanted photophysical processes for improved single-molecule fluorescence imaging. *Biophys. J.* **96**, 2371–2381 (2009).
61. Aitken, C. E., Marshall, R. A. & Puglisi, J. D. An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments. *Biophys. J.* **94**, 1826–1835 (2008).
62. Richards, F. M. & Kundrot, C. E. Identification of structural motifs from protein coordinate data: secondary structure and first-level supersecondary structure. *Proteins* **3**, 71–84 (1988).
63. Xiang, Z., Soto, C. S. & Honig, B. Evaluating conformational free energies: the colony energy and its application to the problem of loop prediction. *Proc. Natl Acad. Sci. USA* **99**, 7432–7437 (2002).
64. Xiang, Z. & Honig, B. Extending the accuracy limits of prediction for side-chain conformations. *J. Mol. Biol.* **311**, 421–430 (2001).
65. Kirschner, K. N. *et al.* GLYCAM06: a generalizable biomolecular force field. *Carbohydrates. J. Comput. Chem.* **29**, 622–655 (2008).
66. Cornell, W. D. *et al.* A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **117**, 5179–5197 (1995).
67. Tomaras, G. D. *et al.* Initial B-cell responses to transmitted human immunodeficiency virus type 1: virion-binding immunoglobulin M (IgM) and IgG antibodies followed by plasma anti-gp41 antibodies with ineffective control of initial viremia. *J. Virol.* **82**, 12449–12463 (2008).
68. Tomaras, G. D. *et al.* Polyclonal B cell responses to conserved neutralization epitopes in a subset of HIV-1-infected individuals. *J. Virol.* **85**, 11502–11519 (2011).
69. Li, M. *et al.* Human immunodeficiency virus type 1 env clones from acute and early subtype b infections for standardized assessments of vaccine-elicited neutralizing antibodies. *J. Virol.* **79**, 10108–10125 (2005).
70. Zhou, T. *et al.* Structural basis for broad and potent neutralization of HIV-1 by antibody VRC01. *Science* **329**, 811–817 (2010).
71. Zhou, T. *et al.* Structural definition of a conserved neutralization epitope on HIV-1 gp120. *Nature* **445**, 732–737 (2007).
72. Scharf, L. *et al.* Antibody 8ANC195 reveals a site of broad vulnerability on the HIV-1 envelope spike. *Cell Rep.* **7**, 785–795 (2014).
73. Calarese, D. A. *et al.* Antibody domain exchange is an immunological solution to carbohydrate cluster recognition. *Science* **300**, 2065–2071 (2003).
74. Xu, R. *et al.* Structural basis of preexisting immunity to the 2009 H1N1 pandemic influenza virus. *Science* **328**, 357–360 (2010).
75. Ekiert, D. C. *et al.* Cross-neutralization of influenza A viruses mediated by a single antibody loop. *Nature* **489**, 526–532 (2012).
76. Sui, J. *et al.* Structural and functional bases for broad-spectrum neutralization of avian and human influenza A viruses. *Nature Struct. Mol. Biol.* **16**, 265–273 (2009).
77. Friesen, R. H. *et al.* A common solution to group 2 influenza virus neutralization. *Proc. Natl Acad. Sci. USA* **111**, 445–450 (2014).
78. McLellan, J. S. *et al.* Structural basis of respiratory syncytial virus neutralization by motavizumab. *Nature Struct. Mol. Biol.* **17**, 248–250 (2010).
79. McLellan, J. S. *et al.* Structure of a major antigenic site on the respiratory syncytial virus fusion glycoprotein in complex with neutralizing antibody 101F. *J. Virol.* **84**, 12236–12244 (2010).
80. Mann, H. B. W. & Donald, R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Stat.* **18**, 50–60 (1947).
81. The PyMOL Molecular Graphics System. (DeLano Scientific, San Carlos, California, 2002).
82. Weis, W. I., Brunger, A. T., Skehel, J. J. & Wiley, D. C. Refinement of the influenza virus hemagglutinin by simulated annealing. *J. Mol. Biol.* **212**, 737–761 (1990).
83. Chen, J., Skehel, J. J. & Wiley, D. C. N- and C-terminal residues combine in the fusion-pH influenza hemagglutinin HA(2) subunit to form an N cap that

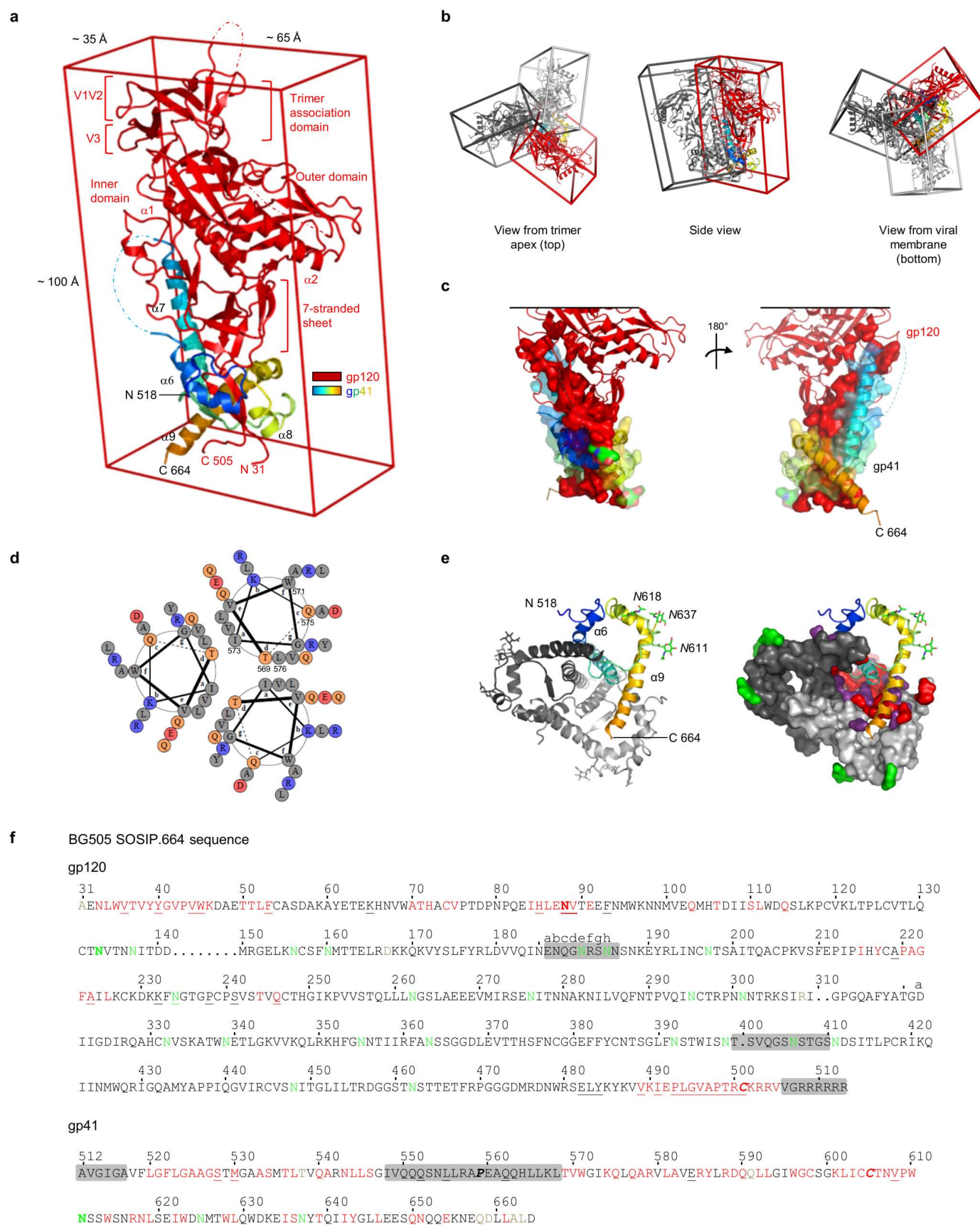
- terminates the triple-stranded coiled coil. *Proc. Natl Acad. Sci. USA* **96**, 8967–8972 (1999).
84. Malashkevich, V. N. *et al.* Core structure of the envelope glycoprotein GP2 from Ebola virus at 1.9-Å resolution. *Proc. Natl Acad. Sci. USA* **96**, 2662–2667 (1999).
  85. Lin, Y. P. *et al.* Evolution of the receptor binding properties of the influenza A(H3N2) hemagglutinin. *Proc. Natl Acad. Sci. USA* **109**, 21474–21479 (2012).
  86. Chen, L. *et al.* Structural basis of immune evasion at the site of CD4 attachment on HIV-1 gp120. *Science* **326**, 1123–1127 (2009).
  87. Rini, J. M. *et al.* Crystal structure of a human immunodeficiency virus type 1 neutralizing antibody, 50.1, in complex with its V3 loop peptide antigen. *Proc. Natl Acad. Sci. USA* **90**, 6325–6329 (1993).
  88. Stanfield, R. *et al.* Dual conformations for the HIV-1 gp120 V3 loop in complexes with different neutralizing fabs. *Structure* **7**, 131–142 (1999).
  89. Tugarinov, V., Zvi, A., Levy, R. & Anglister, J. A *cis* proline turn linking two  $\beta$ -hairpin strands in the solution structure of an antibody-bound HIV-1<sub>IIIIB</sub> V3 peptide. *Nature Struct. Biol.* **6**, 331–335 (1999).
  90. Ofek, G. *et al.* Structure and mechanistic analysis of the anti-human immunodeficiency virus type 1 antibody 2F5 in complex with its gp41 epitope. *J. Virol.* **78**, 10724–10737 (2004).
  91. Cardoso, R. M. *et al.* Broadly neutralizing anti-HIV antibody 4E10 recognizes a helical conformation of a highly conserved fusion-associated motif in gp41. *Immunity* **22**, 163–173 (2005).
  92. Luftig, M. A. *et al.* Structural basis for HIV-1 neutralization by a gp41 fusion intermediate-directed antibody. *Nature Struct. Mol. Biol.* **13**, 740–747 (2006).
  93. Cardoso, R. M. *et al.* Structural basis of enhanced binding of extended and helically constrained peptide epitopes of the broadly neutralizing HIV-1 antibody 4E10. *J. Mol. Biol.* **365**, 1533–1544 (2007).
  94. Grigoryan, G. & Keating, A. E. Structural specificity in coiled-coil interactions. *Curr. Opin. Struct. Biol.* **18**, 477–483 (2008).
  95. Helseth, E., Olshevsky, U., Furman, C. & Sodroski, J. Human immunodeficiency virus type 1 gp120 envelope glycoprotein regions important for association with the gp41 transmembrane glycoprotein. *J. Virol.* **65**, 2119–2123 (1991).
  96. Thali, M., Furman, C., Helseth, E., Repke, H. & Sodroski, J. Lack of correlation between soluble CD4-induced shedding of the human immunodeficiency virus type 1 exterior envelope glycoprotein and subsequent membrane fusion events. *J. Virol.* **66**, 5516–5524 (1992).
  97. Cao, J. *et al.* Effects of amino acid changes in the extracellular domain of the human immunodeficiency virus type 1 gp41 envelope glycoprotein. *J. Virol.* **67**, 2747–2755 (1993).
  98. Leavitt, M., Park, E. J., Sidorov, I. A., Dimitrov, D. S. & Quinnan, G. V. Jr. Concordant modulation of neutralization resistance and high infectivity of the primary human immunodeficiency virus type 1 MN strain and definition of a potential gp41 binding site in gp120. *J. Virol.* **77**, 560–570 (2003).
  99. Yang, X., Mahony, E., Holm, G. H., Kassa, A. & Sodroski, J. Role of the gp120 inner domain beta-sandwich in the interaction between the human immunodeficiency virus envelope glycoprotein subunits. *Virology* **313**, 117–125 (2003).
  100. Sen, J., Jacobs, A. & Caffrey, M. Role of the HIV gp120 conserved domain 5 in processing and viral entry. *Biochemistry* **47**, 7788–7795 (2008).
  101. Wang, J., Sen, J., Rong, L. & Caffrey, M. Role of the HIV gp120 conserved domain 1 in processing and viral entry. *J. Biol. Chem.* **283**, 32644–32649 (2008).
  102. Lawless, M. K. *et al.* HIV-1 membrane fusion mechanism: structural studies of the interactions between biologically-active peptides from gp41. *Biochemistry* **35**, 13697–13708 (1996).
  103. Chen, C. H., Matthews, T. J., McDanal, C. B., Bolognesi, D. P. & Greenberg, M. L. A molecular clasp in the human immunodeficiency virus (HIV) type 1 TM protein determines the anti-HIV activity of gp41 derivatives: implication for viral fusion. *J. Virol.* **69**, 3771–3777 (1995).
  104. Root, M. J., Kay, M. S. & Kim, P. S. Protein design of an HIV-1 entry inhibitor. *Science* **291**, 884–888 (2001).
  105. Gustchina, E. *et al.* Structural basis of HIV-1 neutralization by affinity matured Fabs directed against the internal trimeric coiled-coil of gp41. *PLoS Pathog.* **6**, e1001182 (2010).
  106. Sabin, C. *et al.* Crystal structure and size-dependent neutralization properties of HK20, a human monoclonal antibody binding to the highly conserved heptad repeat 1 of gp41. *PLoS Pathog.* **6**, e1001195 (2010).
  107. Blattner, C. *et al.* Structural delineation of a quaternary, cleavage-dependent epitope at the gp41-gp120 interface on intact HIV-1 Env trimers. *Immunity* **40**, 669–680 (2014).
  108. Yasmeen, A. *et al.* Differential binding of neutralizing and non-neutralizing antibodies to native-like soluble HIV-1 Env trimers, uncleaved Env proteins, and monomeric subunits. *Retrovirology* **11**, 41 (2014).
  109. Falkowska, E. *et al.* Broadly neutralizing HIV antibodies define a glycan-dependent epitope on the prefusion conformation of gp41 on cleaved envelope trimers. *Immunity* **40**, 657–668 (2014).
  110. Thali, M. *et al.* Characterization of conserved human immunodeficiency virus type 1 gp120 neutralization epitopes exposed upon gp120-CD4 binding. *J. Virol.* **67**, 3978–3988 (1993).
  111. Guttman, M. *et al.* CD4-induced activation in a soluble HIV-1 Env trimer. *Structure* **22**, 974–984 (2014).
  112. Scheid, J. F. *et al.* Sequence and structural convergence of broad and potent HIV antibodies that mimic CD4 binding. *Science* **333**, 1633–1637 (2011).
  113. Walker, L. M. *et al.* Broad and potent neutralizing antibodies from an African donor reveal a new HIV-1 vaccine target. *Science* **326**, 285–289 (2009).
  114. Stanfield, R. L. & Wilson, I. A. Structural studies of human HIV-1 V3 antibodies. *Hum. Antibodies* **14**, 73–80 (2005).
  115. Rizzuto, C. D. *et al.* A conserved HIV gp120 glycoprotein structure involved in chemokine receptor binding. *Science* **280**, 1949–1953 (1998).
  116. Guan, Y. *et al.* Diverse specificity and effector function among human antibodies to HIV-1 envelope glycoprotein epitopes exposed by CD4 binding. *Proc. Natl Acad. Sci. USA* **110**, E69–E78 (2013).
  117. Gorny, M. K., VanCott, T. C., Williams, C., Revesz, K. & Zolla-Pazner, S. Effects of oligomerization on the epitopes of the human immunodeficiency virus type 1 envelope glycoproteins. *Virology* **267**, 220–228 (2000).
  118. Yuan, W. *et al.* Oligomer-specific conformations of the human immunodeficiency virus (HIV-1) gp41 envelope glycoprotein ectodomain recognized by human monoclonal antibodies. *AIDS Res. Hum. Retroviruses* **25**, 319–328 (2009).
  119. Moore, P. L. *et al.* Nature of nonfunctional envelope proteins on the surface of human immunodeficiency virus type 1. *J. Virol.* **80**, 2515–2528 (2006).
  120. Frey, G. *et al.* Distinct conformational states of HIV-1 gp41 are recognized by neutralizing and non-neutralizing antibodies. *Nature Struct. Mol. Biol.* **17**, 1486–1491 (2010).
  121. Miller, M. D. *et al.* A human monoclonal antibody neutralizes diverse HIV-1 isolates by binding a critical gp41 epitope. *Proc. Natl Acad. Sci. USA* **102**, 14759–14764 (2005).
  122. Chen, J. *et al.* Mechanism of HIV-1 neutralization by antibodies targeting a membrane-proximal region of gp41. *J. Virol.* **88**, 1249–1258 (2014).
  123. Frey, G. *et al.* A fusion-intermediate state of HIV-1 gp41 targeted by broadly neutralizing antibodies. *Proc. Natl Acad. Sci. USA* **105**, 3739–3744 (2008).
  124. Nicely, N. I. *et al.* Crystal structure of a non-neutralizing antibody to the HIV-1 gp41 membrane-proximal external region. *Nature Struct. Mol. Biol.* **17**, 1492–1494 (2010).
  125. Huang, J. *et al.* Broad and potent neutralization of HIV-1 by a gp41-specific human antibody. *Nature* **491**, 406–412 (2012).
  126. Chakrabarti, B. K. *et al.* HIV type 1 Env precursor cleavage state affects recognition by both neutralizing and nonneutralizing gp41 antibodies. *AIDS Res. Hum. Retroviruses* **27**, 877–887 (2011).
  127. Ruprecht, C. R. *et al.* MPEP-specific antibodies induce gp120 shedding and irreversibly neutralize HIV-1. *J. Exp. Med.* **208**, 439–454 (2011).





**Extended Data Figure 1 | Antibody-mediated crystallization and antibody-induced conformation.** **a**, Atomic-level structures for HIV-1 Env regions determined in complex with HIV-1-neutralizing antibodies. Neutralizing antibodies generally recognize the pre-fusion conformation of HIV-1 Env. Structures highlighted here display a cumulative sum total of pre-fusion HIV-1 Env structural information. Env residues are numbered according to standard HX numbering (from PDBs). One structure, for antibody D5 (blue), is in the post-fusion gp41 conformation, and is not included in the sum total. Regions of other structures (purple) did not define sequence register, and were also not included in the sum total. References listed in this figure are cited elsewhere in the manuscript, except for Rini *et al.* (1993)<sup>87</sup>, Stanfield *et al.* (1999)<sup>88,89</sup>, Ofek *et al.* (2004)<sup>90</sup>, Cardoso *et al.* (2005)<sup>91</sup>, Luftig *et al.* (2006)<sup>92</sup>, Cardoso *et al.* (2007)<sup>93</sup>. **b**, Antibody-induced conformation of HIV-1 Env in the context of infectious JR-FL virions as assessed by smFRET. HIV-1<sub>JR-FL</sub> gp160 was labelled with fluorescent dyes in variable regions, V1 and V4, at positions that did not interfere with Env function (see Methods), and virus was surface-immobilized for imaging via total internal reflection fluorescence

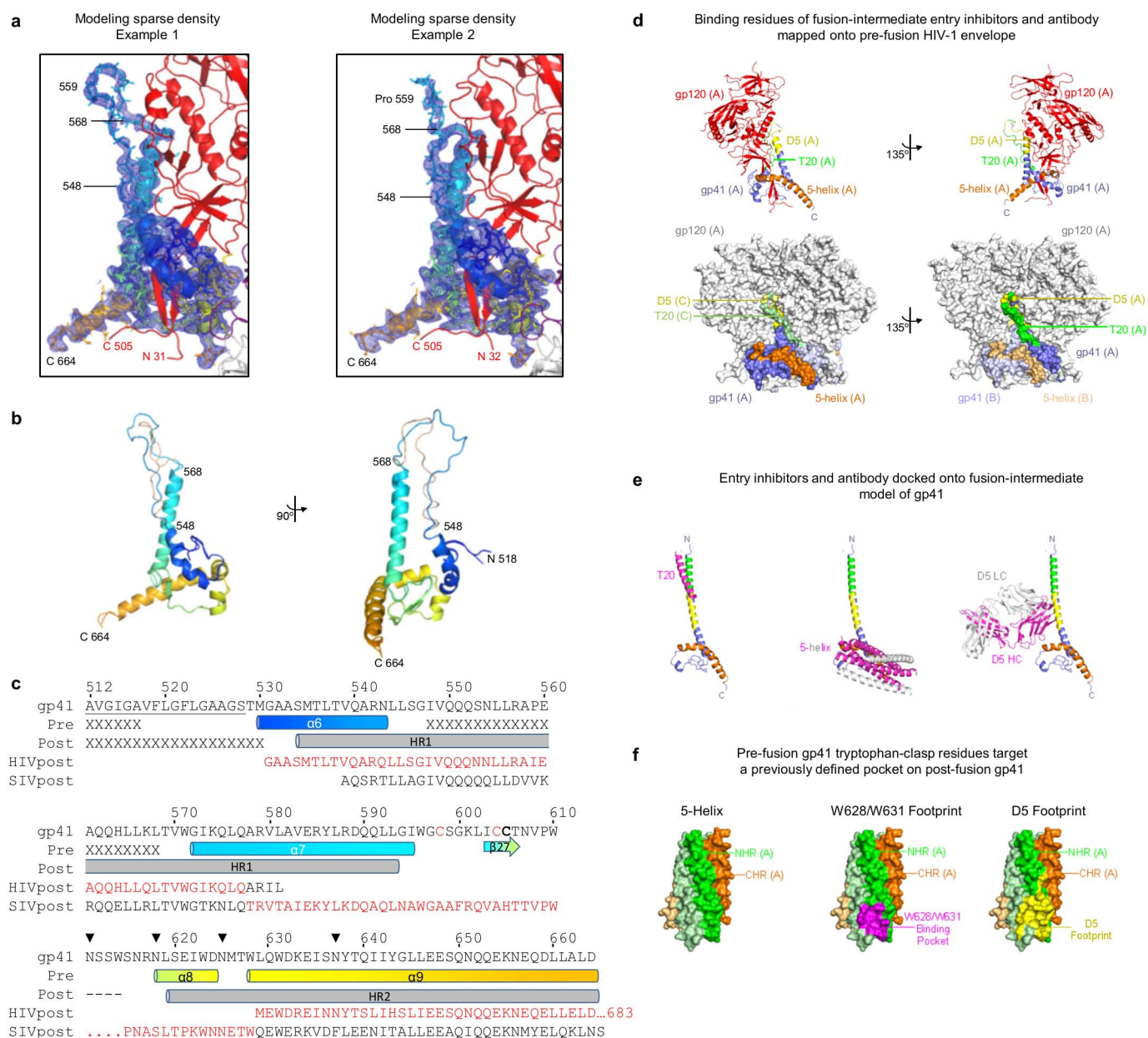
microscopy<sup>18</sup>. smFRET trajectories were compiled into histograms for the HIV-1<sub>JR-FL</sub> Env trimer, either unliganded or after pre-incubation for 30 min with 0.1 mg ml<sup>-1</sup> PGT122, 35O22, or both PGT122 and 35O22 before imaging. Resultant Env conformational landscapes could be deconvoluted into three gaussian distributions: a low-FRET population that predominated for the pre-fusion mature unliganded state, and intermediate- and high-FRET populations, which predominated in the presence of CD4 receptor and CD4-induced antibody<sup>18</sup>. smFRET trajectories are shown for the unliganded HIV-1<sub>JR-FL</sub> Env trimer as well as in the presence of PGT122, 35O22, and both PGT122 and 35O22. The concordance between conformational ensembles indicates unliganded and PGT122 + 35O22-bound conformation to be similar (Spearman correlation coefficient of 0.988). Interestingly, the presence of just one of the antibodies (PGT122) appeared to reduce the high FRET population, an effect not observed in the presence of both antibodies; this suggests that the antibody-induced stability of a particular state is not solely additive, and that antibodies can both induce a particular conformational state as well as alter the transition dynamics from that state.



**Extended Data Figure 2 | HIV-1 subunit interactions: principle component analysis and interface contacts.** **a**, Minimum-bounding box, generated by principle component analysis, encasing 90% of the HIV-1 Env gp120–gp41 protomer. Each gp120–gp41 blade forms a rectangle of height of  $\sim 100$  Å, width of  $\sim 65$  Å, and thickness of  $\sim 35$  Å. Subunits are displayed in ribbon representation with gp41 coloured rainbow and gp120 coloured and labelled red. As previously visualized<sup>10,11</sup>, the membrane-distal portion of the rectangle is made up of the gp120-outer and -inner domains, with the central 7-stranded  $\beta$ -sandwich of the inner domain occupying the trimer-distal, membrane-proximal portion of gp120. We have now resolved the rest of the spike: the membrane-proximal portion of the rectangle is made up of gp41, with the membrane-distal portion of gp41 closest to the molecular threefold axis occupied by helix  $\alpha 7$  (which corresponds in register to the C-terminal portion of the post-fusion HR1 helix of gp41), and the rest of gp41 folding around N and C termini strands of gp120, which extend over 20 Å towards the viral membrane. Of the four helices,  $\alpha 6$  kinks at residue 537<sub>gp41</sub> and  $\alpha 9$  kinks at residue 637<sub>gp41</sub>; backbone H-bonding is less ideal at residues 663<sub>gp41</sub> and 664<sub>gp41</sub>. **b**, Different views of trimeric protomer association. The protomer association at the membrane-distal trimer apex occurs through the corners of the minimum-bounding box, whereas the association at the membrane-proximal region occurs with substantial interpenetration of the minimum-bounding box; these interaction differences and the protruding nature of the gp120 outer domain result in the overall mushroom shape of the trimer.

**c**, gp120–gp41 interface. Ribbon representation of gp120 (red) and gp41 (rainbow from blue N terminus to orange C terminus), with gp120 residues that interact with gp41 shown in surface representation and gp41 residues that interact with gp120 shown in semi-transparent surface. A complete list of subunit interactions is provided in Supplementary Table 1. Membrane-proximal interactions are further stabilized by hydrophobic interactions, which gp41 makes with the N and C termini of gp120, such as between Trp 35<sub>gp120</sub> and Pro609<sub>gp41</sub> and between Trp 610<sub>gp41</sub> and Pro498<sub>gp120</sub>. **d**, Wheel diagram representation of  $\alpha 7$  coiled-coil in the pre-fusion mature closed conformation of gp41 as generated by DrawCoil 1.0: <http://www.grigoryanlab.org/drawcoil/> (ref. 94). **e**, gp41–trimer interfaces as viewed from the viral membrane in ribbon and surface representation ( $90^\circ$  rotation from Fig. 2c). **f**, BG505 SOSIP.664 sequence with residues identified by mutagenesis<sup>95–101</sup> to be important for gp120/gp41 association underlined. Residues that were found to interact between gp120 and gp41 by examination of the crystal structure are indicated in red (intra-protomer interactions) and in brown (inter-protomer interactions). Sites of *N*-linked glycosylation are shown in green; glycan N88 is shown in red because it is part of the gp120/gp41 interactions; no density was observed for potential *N*-linked glycans at residues 185, 398, 406, 411, 462 and 625. Residues that were disordered in the crystal structure are grey. SOS (A501C/T605C) and IP (I559P) mutations are labelled in bold and italics. Dots indicate residues not present in the BG505 sequence.

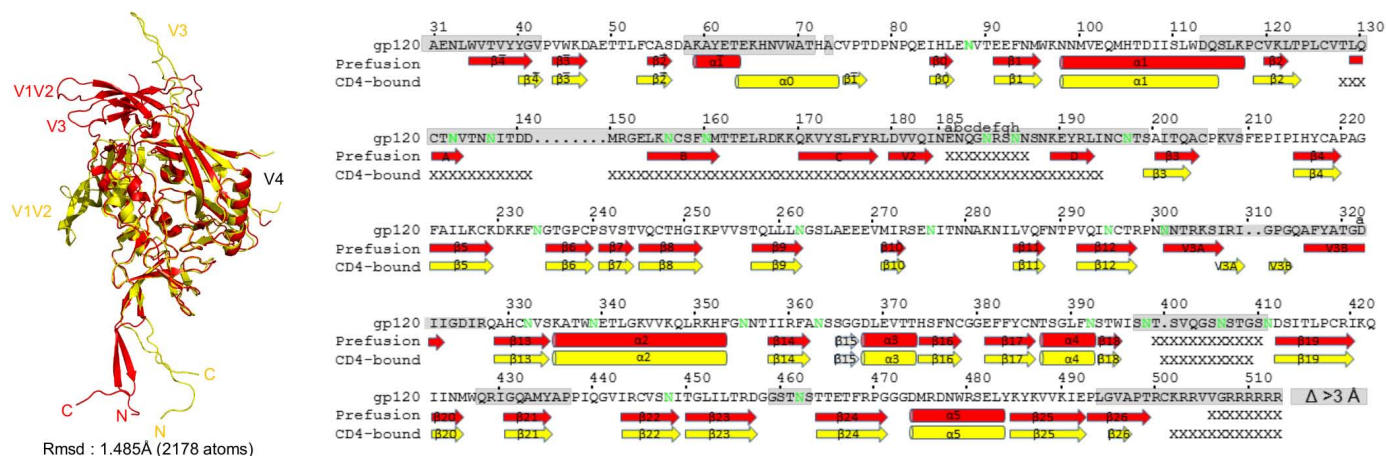




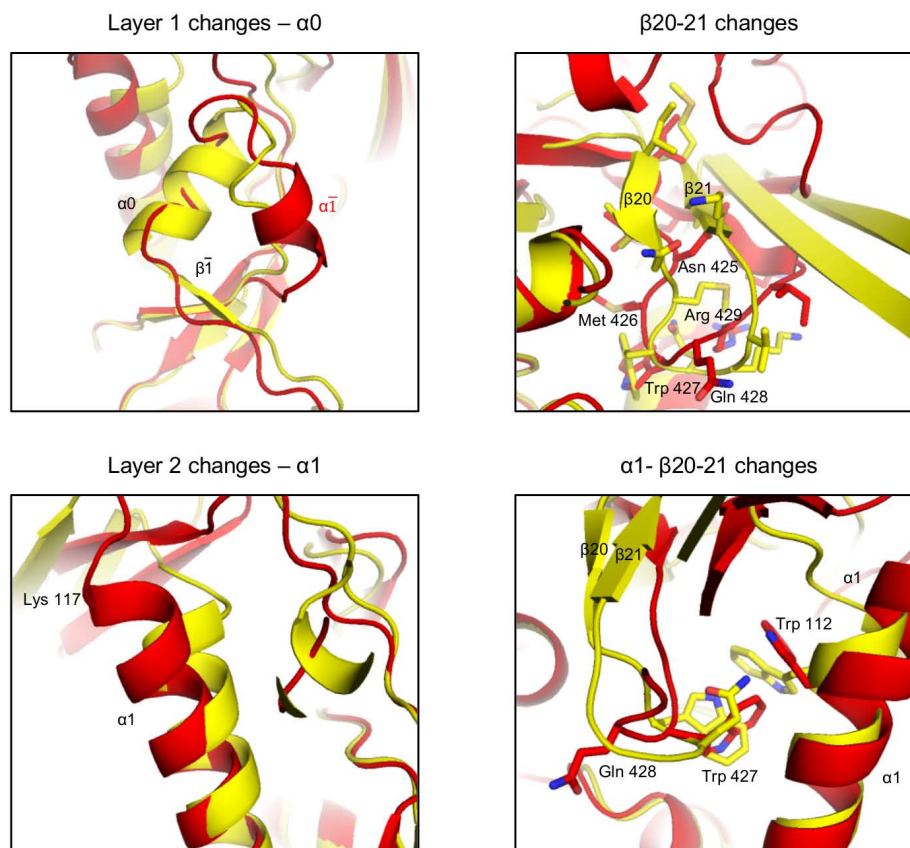
**Extended Data Figure 3 | Modelling of gp41: pre-fusion  $\alpha 6$ -to- $\alpha 7$  density, HIV-1-SIV post-fusion chimaera, and liganded interactions.** **a**, Modelling of gp41 residues 548–568. At low contour, suggestive density is observed that might correspond to the connection between  $\alpha 6$  and  $\alpha 7$  helices. This density appeared to be crystal dependent and might be related to inherent flexibility, functional rearrangements, asymmetry between protomers, or combinations of these factors. To investigate the degree to which a model for this region might be defined, we built and refined two different models for this region: electron density (blue) shown for  $2F_o - F_c$  density at  $1\sigma$  contour; gp41 (rainbow colour from blue to orange) shown in ribbon representation with side chains; gp120 (red) shown in ribbon representation. The location of the I559P mutation is indicated. **b**, The two models from panel **a** are superimposed and shown in perpendicular orientations. **c**, HIV-1-SIV post-fusion chimaera. Sequences of HIV-1 gp41 from pre-fusion structure (BG505 strain, PDB ID 4TVP), post-fusion structure (HIVpost, PDB ID 2X7R<sup>24</sup>) and SIV gp41 post-fusion structure (SIVpost, PDB ID 2EZO<sup>25</sup>) are aligned with secondary structure indicated. Residues that were used to make the post-fusion HIV-1-SIV chimaera used in Fig. 3 are highlighted in red. **d**, Binding residues of representative fusion-intermediate entry inhibitors or antibodies mapped onto the structure

of pre-fusion HIV-1 Env spike<sup>102–104</sup>. Top, ribbon representation of pre-fusion envelope protomer A (gp120 in red and gp41 in blue) at two orientations, with the binding residues of the fusion-intermediate inhibitors 5-helix<sup>104</sup> and T20<sup>102,103</sup> and of monoclonal antibody D5<sup>92</sup> shown in orange, green and yellow, respectively. Bottom, surface representation of the pre-fusion envelope trimer, with inhibitor and antibody binding residues mapped onto the surfaces of all protomers (A, B, C). gp120 is coloured grey and gp41 is coloured in shades of blue, depending on protomer. Binding residues of fusion-intermediate inhibitors 5-helix, T20 and monoclonal antibody D5 are shown in same colour shades as in the top panels. **e**, 5-helix, T20 and D5 Fab (all coloured magenta and grey) docked onto a model of fusion-intermediate gp41 (coloured as in **d**). **f**, A previously defined binding pocket on post-fusion gp41 is recognized by pre-fusion gp41 tryptophan-clasp residues Trp 628 and Trp 631. Shown is a surface representation of gp41 5-helix protein<sup>104</sup> (left, with N-heptad repeat (NHR) helices coloured in shades of green and C-heptad repeat (CHR) helices coloured in shades of orange). The footprint of gp41 tryptophan-clasp residues Trp 628 and Trp 631 is shown in magenta (middle) and that of a representative NHR-specific neutralizing antibody, D5, in yellow<sup>92,105,106</sup> (right).

## a Pre-fusion gp120 states: mature closed versus CD4-bound

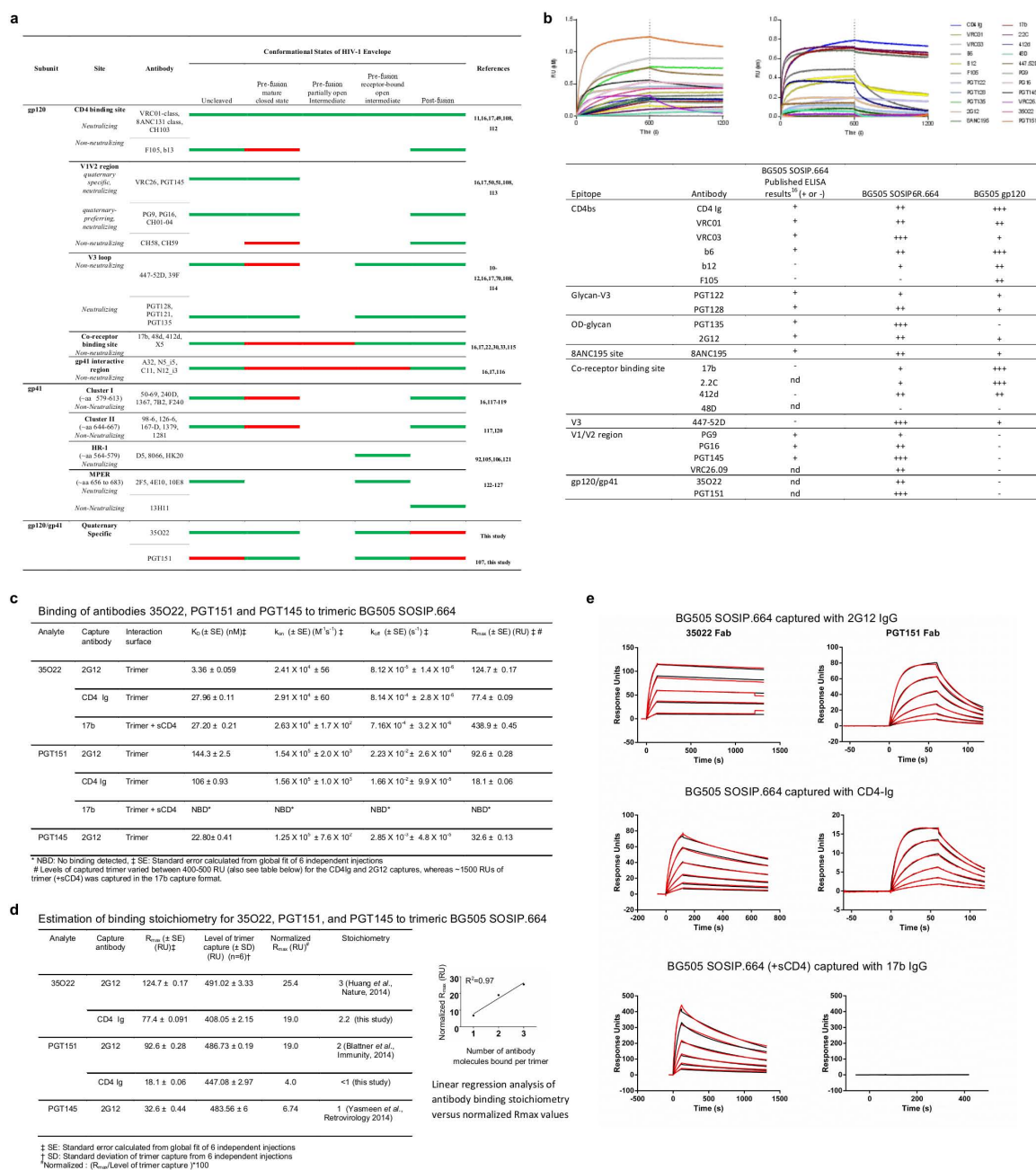


## b



**Extended Data Figure 4 | Conformational changes between pre-fusion mature closed state and CD4-bound state of gp120.** **a**, Overall structure and sequence comparison. gp120 is shown in ribbon representation in pre-fusion mature closed (red) and CD4-bound (yellow, PDB ID 3JWD<sup>22</sup>) conformation. V1V2 (PDB ID 3U2S<sup>51</sup>) and V3 (PDB ID 2B4C<sup>30</sup>) have been modelled onto the CD4-bound conformation. Secondary structure is defined for pre-fusion and CD4-bound conformation on the BG505 sequence, with cylinders representing  $\alpha$ -helix and arrows  $\beta$ -strands. Disordered residues are indicated by X. Residues that move more than 3 Å between the mature closed and the

CD4-bound gp120 conformations are highlighted by grey shadows. Sites of N-linked glycosylation are shown in green. **b**, Details of conformational changes between the mature closed (red) and the CD4-bound conformations (yellow) of gp120 (shown in ribbon). Regions highlighted cover layer 1 with changes at  $\alpha 0$  (we note that density in this region is not well defined), layer 2 with changes at  $\alpha 1$  and  $\beta 20$ -21 rearrangements. All atoms r.m.s.d. are as follows: residues 54<sub>gp120</sub>-74<sub>gp120</sub>, r.m.s.d. = 4.759 Å; residues 98<sub>gp120</sub>-117<sub>gp120</sub>, r.m.s.d. = 0.497 Å; 424<sub>gp120</sub>-436<sub>gp120</sub>, r.m.s.d. = 3.196 Å.





**a**

Probe Radius (R=1.4 Å)		
	Fraction Glycan Covered	Fraction Glycan Free
HIV	0.71	0.29
HIV <sub>241</sub>	0.72	0.28
Flu	0.54	0.46
RSV	0.26	0.74

Probe Radius (R=10.0 Å)		
	Fraction Glycan Covered	Fraction Glycan Free
HIV	0.97	0.03
HIV <sub>241</sub>	0.97	0.03
Flu	0.86	0.14
RSV	0.52	0.48

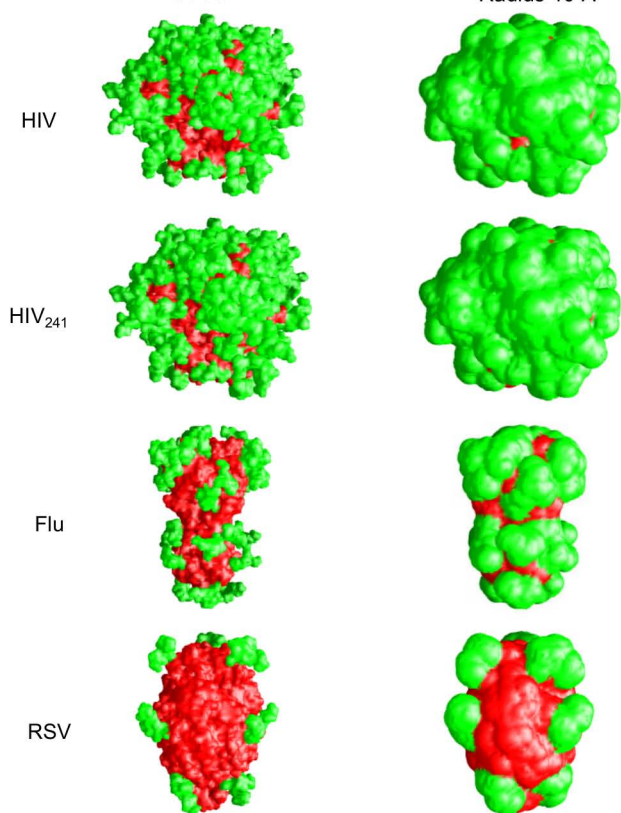
**b**

Probe Radius (R=1.4 Å)		
	Fraction Glycan Covered	Fraction Glycan Free
HIV	0.63	0.37
HIV <sub>241</sub>	0.63	0.37
Flu	0.44	0.56
RSV	0.19	0.81

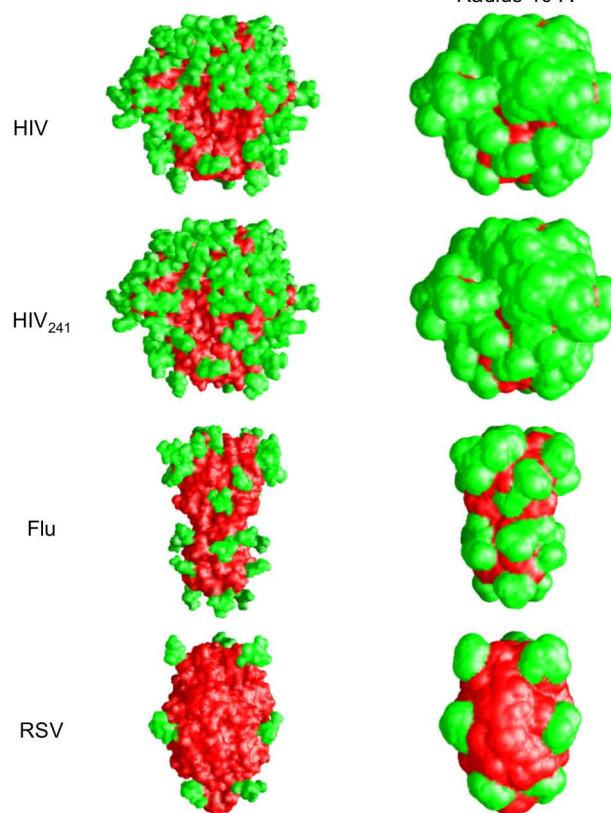
  

Probe Radius (R=10.0 Å)		
	Fraction Glycan Covered	Fraction Glycan Free
HIV	0.94	0.06
HIV <sub>241</sub>	0.95	0.05
Flu	0.77	0.23
RSV	0.42	0.58

**C**

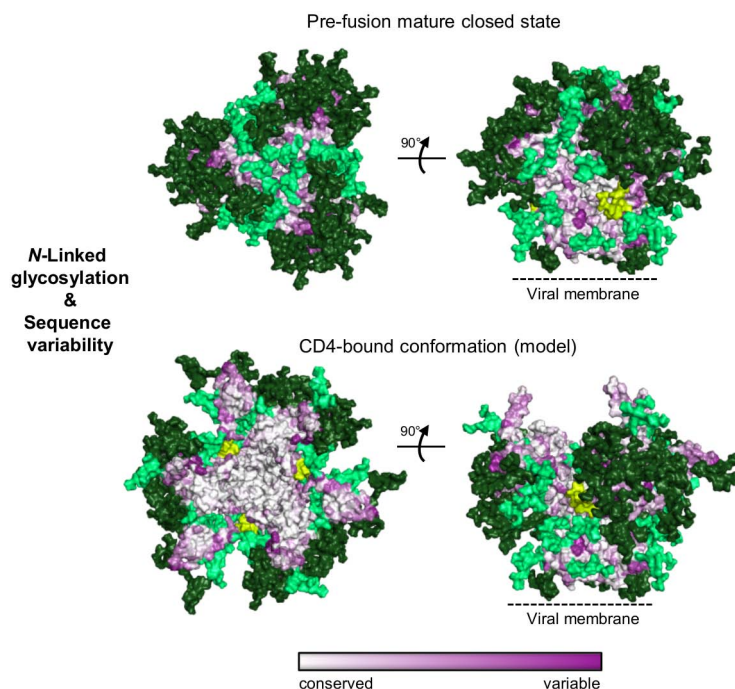


**d**



**Extended Data Figure 6 | N-linked glycan occlusion of type I fusion machines.** The pre-fusion mature closed conformation of HIV-1 Env evades the humoral immune response with a fully assembled glycan shield. Here we calculate and display the solvent-accessible surface of glycan and protein for HIV-1 Env, HIV<sub>241</sub> (which contains an added glycan at position N241), influenza virus haemagglutinin and RSV fusion glycoprotein. Calculations of the percentage coverage of the protein surface were determined for trimeric type I fusion machines based on two probe sizes of 1.4 Å (solvent radius) and 10.0 Å (the estimated steric footprint of an antibody combining region). Surface area calculations were carried out according to Kong *et al.*<sup>79</sup>, and images

were generated using Grasp v1.3<sup>80</sup>. All models were refined using Amber with the GLYCAM force field (see Methods for details). The PDB IDs associated with the glycosylated models are: 4TVP (HIV-1), 2YP7<sup>85</sup> (Flu) and 4JHW<sup>31</sup> (RSV). The strains associated with the PDB IDs are: BG505.SOSIP.664 (HIV-1), H3N2 A/Hong Kong/4443/2005 (Flu) and A/A2/61 (RSV). The solvent-accessible protein surface is shown in red, and N-linked glycans are shown in green. **a**, Estimated Man9 glycan coverage. **b**, Estimated Man5 glycan coverage. **c**, Visualization of Man9 N-linked glycan coverage for two probe radii. **d**, Visualization of Man5 N-linked glycan coverage for two probe radii.

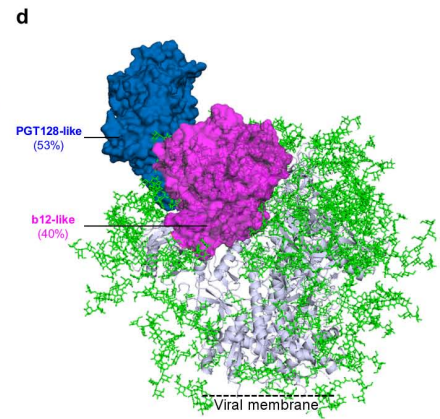
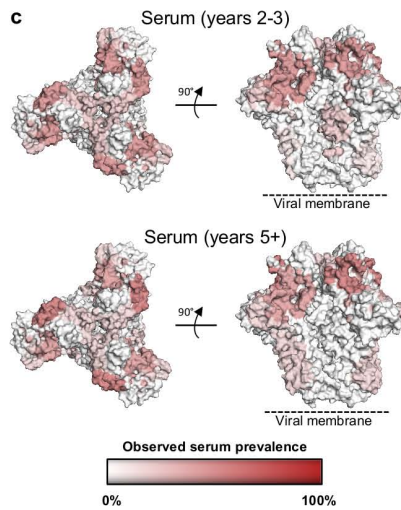
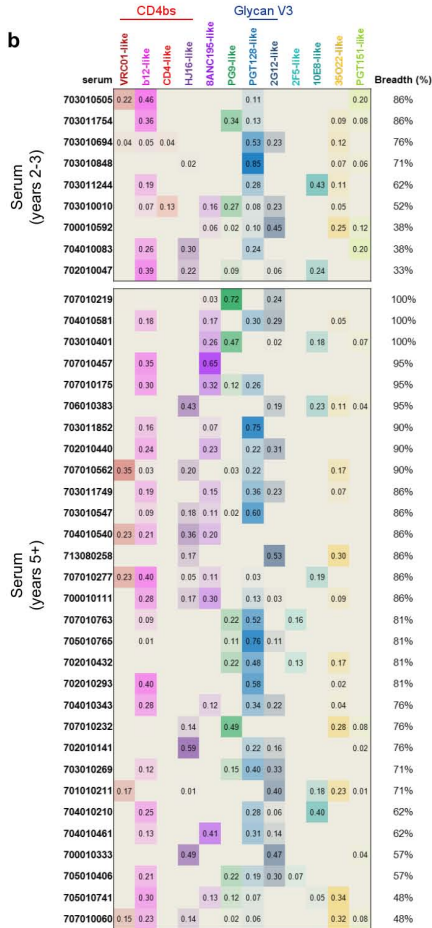


**Extended Data Figure 7 | Glycan shield and sequence variability for HIV-1 pre-fusion mature closed and CD4-bound conformations.** Many conformations of HIV-1 Env divert the immune response. Thus for example, shed gp120 and post-fusion gp41 represent dominant viral antigens; however, these forms of Env are not functional, and antibodies that only target them are not neutralizing. Functional conformations, however, may be significantly shielded from the neutralizing antibody. The CD4-bound conformation of HIV-1 Env, for example, is only functionally present when the viral and target-cell membranes are in close proximity, and the exposed co-receptor binding site (including V3- and CD4-induced epitopes) is spatially occluded from

neutralizing antibody. Here we provide models for the pre-fusion closed state versus the CD4-bound conformation, which display the fully assembled glycan shield and surface Env variability. Env N-linked glycans are depicted in light green (conserved; greater than 90% conservation) or dark green (variable; less than 90% conservation) on the mature closed Env structure and modelled CD4-bound conformation. Env sequence variability is shown from white to purple (conserved to variable). A conserved glycan at residue 241<sub>gp120</sub> not present in the BG505 sequence is shown in yellow-green. As can be seen, the pre-fusion closed state has few glycan-free surfaces, whereas the CD4-bound state exposes substantial glycan-free conserved surface.

**a**

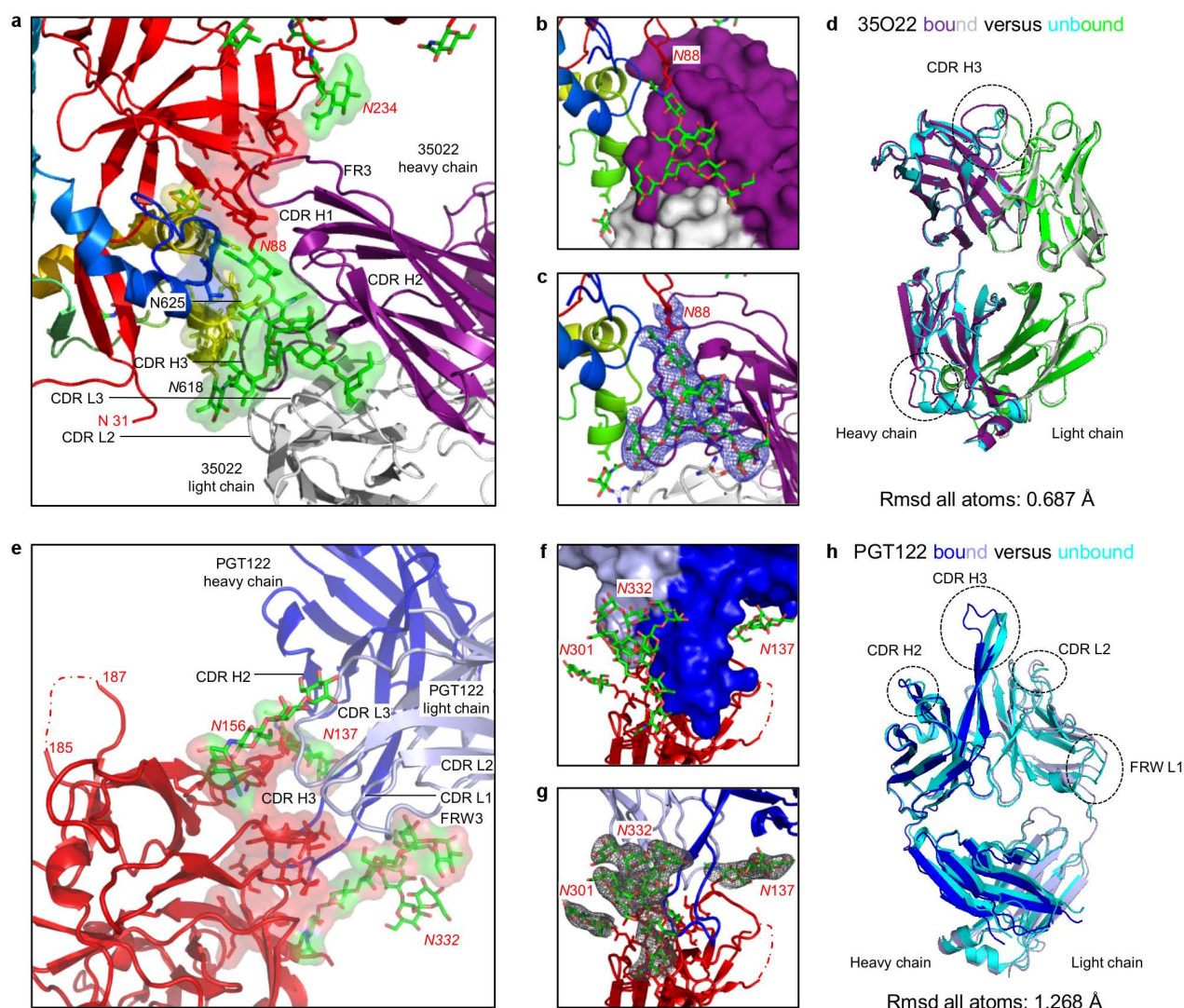
Serum ID	6101.10	Bal.01	BG1168.01	CAAN.A2	DU156.12	DU422.01	JRCSEF.JB	JRFL.JB	KER2018.11	PVO.04	Q168.a2	Q23.17	Q769.h5	RW020.2	THRO.18	TRJO.58	TRO.11	YU2.DG	ZA012.29	ZM106.9	ZM55.28a
703010505	65	376	41	62	304	220	242	1370	123	40	220	180	62	320	<40	<40	61	250	50	56	65
703011754	124	1097	69	280	1535	895	813	335	572	<40	184	1209	349	524	42	<40	<40	455	115	193	79
703010694	158	314	<40	57	150	53	1258	628	40	282	116	435	31	1073	<40	125	144	264	252	90	<40
703010848	240	303	<40	899	144	139	273	474	<40	135	<40	1483	<40	5883	<40	<40	434	122	136.5	440	239
703011244	126	111	40	44	106	152	258	79	44	<40	<40	124	84	296	47	<40	116	40	<40	<40	<40
703010010	<40	283	<40	<40	40	<40	165	<40	<40	53	41	91	44	71	58	<40	67	104	<40	<40	43
700010592	<40	351	<40	63	54	<40	89	<40	<40	105	<40	270	<40	<40	<40	40	68	<40	<40	<40	46
704010083	<40	<40	<40	127	99	68	52	77	<40	<40	<40	<40	<40	102	<40	<40	<40	<40	67	110	<40
702010047	<40	184	<40	<40	40	108	5837	<40	<40	<40	90	<40	<40	51	224	<40	196	<40	<40	<40	<40
707010457	113	152	<40	44	390	467	87	700	146	1304	351	83	613	1591	217	740	684	841	187	170	698
703011852	437	287	<40	439	998	974	553	1245	76	352	87	826	77	7041	<40	91	2652	786	403	301	389
707010219	157	175	187	179	219	142	55991	2203	3421	1835	3640	7177	19080	1172	7584	386	2792	976	219	487	5522
703011749	208	241	<40	87	998	310	5221	3688	<40	1183	56	755	177	2312	60	232	602	174	<40	750	153
704010581	273	1193	69	282	787	337	1415	652	199	1009	135	639	109	2129	267	985	2695	892	178	306	163
703010547	98	121	<40	127	1324	508	101	459	57	226	58	996	<40	4754	<40	57	1912	583	54	481	130
703010401	189	472	188	477	3179	918	1653	876	780	777	1869	992	1193	760	937	1876	1962	490	359	706	1628
707010763	548	857	<40	163	3447	1913	208	1209	3140	300	192	54827	<40	2155	<40	<40	854	207	299	551	146
704010210	246	298	78	<40	3107	400	150	60	<40	211	<40	78	<40	329	78	<40	168	<40	63	<40	<40
707010175	83	81	48	71	400	170	322	244	66	325	118	861	76	1335	182	118	272	144	<40	149	73
703010269	138	639	<40	<40	444	97	2514	117	<40	1041	<40	1160	<40	606	208	<40	414	120	128	72	61
702010440	142	89	57	117	354	154	609	306	42	892	73	88	<40	613	110	275	562	93	<40	87	59
705010741	<40	101	<40	<40	417	140	1171	<40	<40	45	60	94	<40	66	<40	134	<40	<40	<40	50	<40
704010343	218	511	<40	96	297	162	1397	842	<40	875	<40	107	<40	1629	46	594	148	194	<40	252	109
705010765	287	260	<40	517	769	458	1630	658	<40	427	<40	845	<40	2514	57	112	1043	221	424	804	151
702010432	64	279	<40	164	197	76	996	191	1413	61	<40	2108	<40	322	<40	97	153	58	248	955	55
704010461	86	214	<40	<40	549	58	59	74	<40	208	<40	67	<40	125	<40	337	456	130	<40	232	<40
707010562	94	458	<40	109	340	97	9672	1132	<40	81	787	2467	433	3603	41	145	647	360	122	383	174
706010383	292	374	64	1216	198	247	346	123	50	109	2154	94	<40	193	153	255	522	94	50	150	92
704010540	135	193	<40	<40	176	50	131	181	<40	83	281	46	113	319	165	345	205	129	83	219	99
713080258	<40	534	51	82	51	83	171	310	59	81	157	<40	58	51	<40	200	154	53	51	120	58
707010277	80	301	<40	62	667	203	41	425	<40	76	238	177	266	404	231	<40	102	136	68	52	78
702010293	66	504	<40	244	322	218	345	617	<40	48	46	547	<40	1030	108	<40	211	89	63	166	42
700010111	168	785	46	<40	227	194	451	115	<40	175	753	277	<40	1053	107	468	457	399	41	143	58
707010232	<40	270	<40	50	233	144	263	<40	638	52	1754	1023	75	1640	51	43	50	<40	<40	102	48
702010141	122	226	<40	51	78	<40	71	<40	<40	68	760	45	<40	258	52	47	135	60	59	170	116
701010211	55	177	76	54	79	<40	181	159	<40	63	533	135	<40	<40	47	197	78	59	46	<40	<40
700010333	54	410	<40	75	<40	<40	<40	147	<40	66	696	<40	<40	<40	70	53	79	57	<40	71	99
705010406	<40	79	<40	<40	54	175	416	202	84	43	<40	146	<40	80	<40	<40	93	<40	43	82	<40
707010060	<40	123	<40	<40	76	49	149	51	<40	<40	181	125	<40	151	<40	67	<40	<40	<40	48	<40



**Extended Data Figure 8 | Prevalence of neutralizing responses identified serologically from cohorts from 2–3 years and 5+ years post infection.**  
**a**, Serum neutralization on 21-strain virus panel. ID<sub>50</sub>s (reciprocal dilution at which 50% of the virus is neutralized) are shown for serum (rows) titrated against HIV-1 viral strains (columns). **b**, For each serum, the predicted neutralization prevalence for each of 12 antibody specificities is shown based on neutralization of 21 diverse HIV-1 strains. Values of at least 0.2 were considered positive and counted toward the overall cohort prevalence

percentages in Fig. 6c. **c**, Prevalence of antibody specificities onto the HIV-1 Env, coloured as indicated in the bar graph. **d**, The antibody specificities for high serum prevalence in the 5+ years cohort are depicted by Fabs of representative antibodies binding the BG505 SOSIP.664 Env trimer, shown in grey ribbon representation, with glycans as green sticks. Note that while prevalence between the two cohorts showed good correspondence, there were notable differences, for example, between PGT151 at 2–3 years and 5+ years in this study as well as between the cohorts analysed here and in ref. 13.

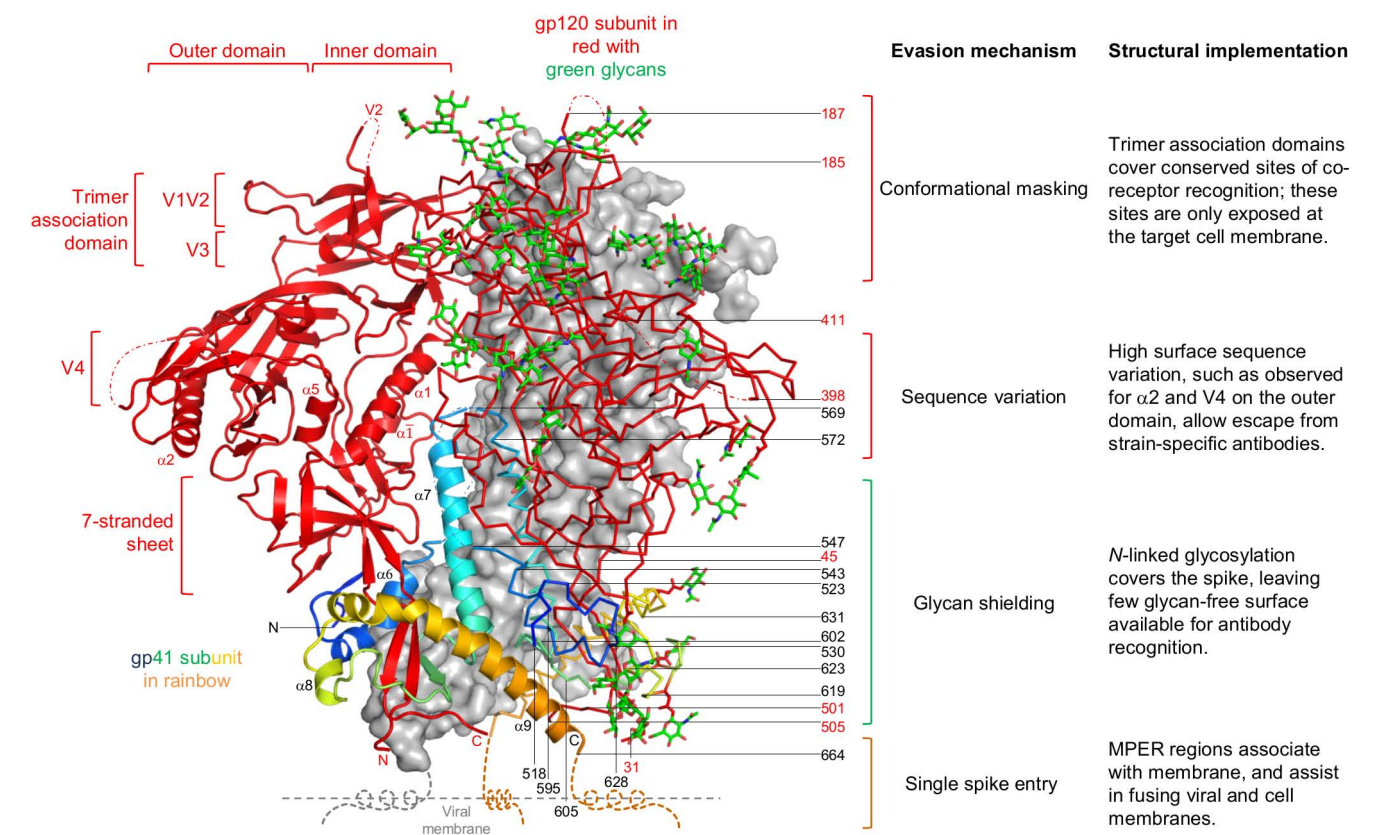




#### Extended Data Figure 9 | Antibodies 35O22 and PGT122: interface with HIV-1 Env and comparison of bound and unbound Fab conformations.

Despite the substantial immune evasion protecting the mature unliganded state from humoral recognition, after several years of infection, the human immune system does generate broadly neutralizing antibodies. 35O22 and PGT122 are two of these antibodies, which neutralize 62% and 65% of HIV-1 isolates at a median  $IC_{50}$  (half maximal inhibitory concentration) of 0.033 and  $0.05 \mu\text{g ml}^{-1}$ , respectively<sup>13,12</sup>. Here we provide additional details on 35O22 and PGT122 recognition. **a**, 35O22 Fab is shown in ribbon representation (purple (heavy chain) and white (light chain)). The gp120 subunit is shown in red, the gp41 subunit in rainbow (from blue N terminus to orange C terminus), and glycans in green sticks. Complementarity determining regions (CDRs) are labelled, and interactive HIV-1 Env residues highlighted in semi-transparent surface representation. At the membrane-distal surface of 35O22, an extended framework 3 region (FW3) of the heavy chain (resulting from an insertion of 8 residues) interacts with strand  $\beta 1$  of the 7-stranded inner domain sandwich of gp120. The heavy chain-CDRs form extensive contacts with the N-linked glycan extending from residue 88<sub>gp120</sub>. In addition to glycan contacts, the CDR H3 of 35O22 interacts with the  $\alpha 9$  helix of gp41. Helix  $\alpha 9$  interactions are also made by the FW3 of the light chain (a complete list of contacts is provided in Supplementary Table 3). Overall, 35O22 buries  $1,105 \text{ \AA}^2$  solvent surface on gp120 (including  $793 \text{ \AA}^2$  with the Asn 88<sub>gp120</sub> glycan) and  $594 \text{ \AA}^2$  solvent surface on gp41 (including  $127 \text{ \AA}^2$  with the Asn 618<sub>gp41</sub> glycan). Despite residue 625<sub>gp41</sub> being part of the glycan sequon 'NMT', no glycan is observed; indeed, the side-chain amide of residue 625<sub>gp41</sub> hydrogen bonds with the side-chain oxygen of Tyr 32 in the 35O22 heavy chain, and the presence of an N-linked glycan at residue 625<sub>gp41</sub> is difficult to reconcile with 35O22 recognition. **b**, Same colours as **a**, with 35O22 Fab shown in surface representation. **c**, Same colours as **a**, with  $2F_o - F_c$  at  $1\sigma$  contour (blue density) shown around glycan 88 of gp120. Antibody 35O22 employs a novel mechanism of glycan-protein

recognition, combining a protruding FW3 with CDR H1, H2 and H3 to form a 'bowl' that holds glycan. FW3 and CDR H3 provide the top edges of the bowl and interact with the protein surface of gp120, whereas CDR H1 and H2 are recessed and hold/recognize glycan. This structural mechanism of recognition contrasts with the extended CDR H3-draping glycan observed with other antibodies that penetrate the glycan shield such as PG9<sup>51</sup> and PGT128<sup>78</sup>. **d**, Unbound and HIV-1 Env-bound 35O22 Fabs were superimposed, and ribbon representations and r.m.s.d.s are displayed. Unbound 35O22 Fab is coloured cyan (heavy chain) and green (light chain), and bound 35O22 Fab is coloured deep purple (heavy chain) and white (light chain). Regions that showed conformational changes are highlighted with black dotted lines. We note that in the 35O22-bound conformation, density is poor and/or sparse for the Fc portion of the Fab. **e**, PGT122 interface details. Ribbon representation of PGT122 Fab in blue (heavy chain) and light blue (light chain) interacting with one gp120 subunit, shown in red with glycans in green sticks. CDRs are labelled, and interactive HIV-1 Env residues highlighted in surface representation. Primary contacts between antibody PGT122 and N-linked glycan involve N137 and N332, with minor contact with N156. Although portions of glycan N301 can be observed in the electron density, no direct contacts with PGT122 are observed; a complete list of contacts between PGT122 and BG505 SOSIP.664 is provided in Supplementary Table 4. **f**, Same colours as **e**, with PGT122 Fab shown in surface representation. **g**, Same colours as **e**, with  $2F_o - F_c$  at  $1\sigma$  contour (grey density) shown around glycan 332 of gp120. **h**, Comparison of bound and unbound PGT122 Fab conformations. Unbound and HIV-1 Env-bound Fabs were superimposed, and ribbon representations and r.m.s.d.s are displayed. Unbound PGT122 Fab is coloured cyan, and bound PGT122 Fab blue (heavy chain) and light blue (light chain). Regions which showed conformational changes are highlighted with black dotted lines.



**Extended Data Figure 10 | Structural implementation of HIV-1 molecular trickery.** The pre-fusion HIV-1 Env trimer (left) is displayed with evasion mechanisms and their structural implementation (right). The gp120 subunit is shown in red, the gp41 subunit in rainbow (from blue N terminus to orange C terminus), and crystallographically defined glycans in green. One protomer is shown with Cα trace and glycans in stick representation; a second protomer

is shown in ribbon representation with secondary structure elements labelled; and the third protomer is shown in light grey surface. The MPER region for each protomer is shown as a stylized helix associated with the viral membrane. The location of secondary structural elements, termini, and residues called in the text has been labelled (red font for gp120 and black font for gp41).

Extended Data Table 1 | Data collection and refinement statistics

BG505 SOSIP.664 with PGT122 and 35O22	
<b>Data collection</b>	
Space group	P6 <sub>3</sub>
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	128.89, 128.89, 313.42
$\alpha$ , $\beta$ , $\gamma$ (°)	90.0, 90.0, 120.0
Resolution (Å)	50-3.10 (3.68-3.49, 3.49-3.34, 3.34-3.21, 3.21-3.10)*
<i>R</i> <sub>sym</sub> or <i>R</i> <sub>merge</sub>	11.6 (44.9, 52.7, 68.3, 75.5)
<i>I</i> / $\sigma$ <i>I</i>	11.3 (2.22, 1.95, 1.43, 1.11)
Completeness (%)	73.4 (57.2, 40.6, 33.9, 27.8)
Redundancy	3.8 (3.3, 3.4, 3.3, 3.0)
<b>Refinement</b>	
Resolution (Å)	40.739-3.098 (3.4893-3.3375; 3.3375-3.2091; 3.2091-3.0984)
No. reflections	29,353
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	0.2135/0.2480 (0.3056/0.3484; 0.3348/0.3315; 0.4039/0.4056)
No. atoms	12189
Protein	11345
Ligand (glycans)	757
Sulfate ion	10
Water	37
B-factors (Å <sup>2</sup> )	109.50
Protein	109.30
Ligand (glycans)	113.34
Sulfate ion	140.95
Water	40.20
R.m.s deviations	
Bond lengths (Å)	0.002
Bond angles (°)	0.575
Ramachadran Favored %	92.66
Ramachadran Allowed %	99.03
MolProbity all-atoms clashscore	5.36
MolProbity score	1.80
<b>PDB ID</b>	<b>4TVP</b>

\* Highest resolution shells are shown in parentheses.



Extended Data Table 2 | Modelling parameters for gp120 and gp41 rearrangements

	Pre-fusion mature closed state	Pre-fusion partially open intermediate	Pre-fusion receptor-bound open intermediate	Post-fusion
gp120	Crystal structure (4TVP)	Crystal structure (4TVP)	Crystal structure of core (3JWD) with modeled V3 (3HI1) and V1V2 (3U4E)	Crystal structure of core (3JWD) with modeled V3 (3HI1) and V1V2 (3U4E)
V1V2	Native	Rotated 6°	Rotated to align with bridging sheet	Rotated to align with bridging sheet
V3	Native	Rotated 6°	Protruding towards target cell	Protruding towards target cell
Core	Native	Rotated 6°	Rotated 50°	Rotated 50°
N+C-term	Native	Native	Unknown	Rotated 45°
gp41	Crystal structure (4TVP)	Crystal structure (4TVP) with modeled $\alpha 7$	Crystal structure (4TVP) with modeled $\alpha 7$ and $\alpha 6$ removed	Crystal structure (chimaera of 2X7R and 2EZO)
$\alpha 6$	Native	Disassembling to $\alpha 7$	Disassembled	Extended to post-fusion HR1
$\alpha 7$	Native	Extending	Extended with fusion peptide	Extended to post-fusion HR1
$\alpha 8$	Native	Native	Native	Extended to post-fusion HR2
$\alpha 9$	Native	Native	Native	Extended to post-fusion HR2

To provide reference frames for the various pre-fusion conformational states, we extracted Env component of SOSIP bound by VRC03<sup>37</sup> and by VRC-PG04<sup>11</sup> (also called PGV04), and aligned the resultant maps with the CD4-bound conformation of trimeric BAL<sup>28</sup> (using Chimera <http://www.cgl.ucsf.edu/chimera/>). Once maps were aligned, gp120 and gp41 models were fit to each of the maps as defined in the table above in black text to the right of labels 'gp120' and 'gp41'. In addition to rigid-body fits of crystal structures, specific regions of gp120 and gp41 were modelled. These are defined in the table above in red text after different portions of gp120 and gp41 relative to the pre-fusion mature closed state. The 35022-/PGT122-bound BG505 SOSIP.664 structure analysed here was found by smFRET to closely resemble the pre-fusion unliganded conformation. It is 'mature', with the C terminus of gp120 cleaved from the N terminus of gp41. We do not know the structure of gp41 in the uncleaved state, but antigenic differences with the mature cleaved state suggest a distinct gp41 conformation; in our pre-fusion HIV-1 Env structure, the observed C terminus of gp120 at residue 505<sub>gp120</sub> and N terminus of gp41 at residue 518<sub>gp41</sub> are 37 Å apart, a distance which cleavage may help the pre-fusion structure to accommodate. Finally, the interactions between V1V2 and V3 at the trimer apex indicate a closed conformation. Altogether, the crystal structure defined here (PDB ID 4TVP) is thus in the pre-fusion mature closed state. Recently, hydrogen-deuterium exchange and oxidative labelling were used by Guttman *et al.*<sup>111</sup> to identify regions in BG505 SOSIP.664 with ordered secondary structure and/or with altered protection upon CD4 binding; these data agree with both structures and modelling presented here. We fit 4TVP without modification to EMDB-5779<sup>10</sup> with density from VRC-PG04 Fabs computationally removed. The pre-fusion partially open intermediate conformation was modelled by a rigid body fitting of gp120 to EMDB-2484<sup>27</sup> with density from VRC03 Fabs computationally removed.  $\alpha 7$  of gp41 was extended into the unoccupied density at the N terminus of the helix using the mature closed structure as a starting model. The pre-fusion receptor-bound intermediate was modelled by fitting the CD4-bound gp120 core crystal structure (PDB ID 3JWD<sup>25</sup>) to the CD4-bound EMDB-5455<sup>28</sup> map. V3 of the crystal structure (PDB ID 3HI1<sup>86</sup>) was aligned to the core and the V1V2 crystal structure (PDB ID 3U4E<sup>51</sup>) was fit to the remaining density.  $\alpha 7$  of gp41 was extended through an alignment with crystal structures of post-fusion gp41 (PDB IDs 2X7R<sup>24</sup>, 2EZO<sup>25</sup>). Post-fusion gp120 is in the same conformation as the pre-fusion receptor-bound intermediate and the post-fusion gp41 structure was derived from an alignment of SIV and HIV post-fusion crystal structures (PDB IDs 2X7R<sup>24</sup>, 2EZO<sup>25</sup>). Note that the fit of the BG505 SOSIP.664 structure defined here to the pre-fusion mature closed state was very good, the fit to the pre-fusion partially open intermediate was good except for the V1V2 region and the membrane-proximal region of gp41, and the fit of the CD4-bound core gp120 to the pre-fusion receptor-bound open intermediate state was similar to the published fit. Overall, we note that these rigid-body fittings only approximate the actual molecular motions between conformations.

# Two families of exocomets in the $\beta$ Pictoris system

F. Kiefer<sup>1,2,3</sup>, A. Lecavelier des Etangs<sup>1,2</sup>, J. Boissier<sup>4</sup>, A. Vidal-Madjar<sup>1,2</sup>, H. Beust<sup>5</sup>, A.-M. Lagrange<sup>5</sup>, G. Hébrard<sup>1,2</sup> & R. Ferlet<sup>1,2</sup>

The young planetary system surrounding the star  $\beta$  Pictoris harbours active minor bodies<sup>1–6</sup>. These asteroids and comets produce a large amount of dust and gas through collisions and evaporation, as happened early in the history of our Solar System<sup>7</sup>. Spectroscopic observations of  $\beta$  Pictoris reveal a high rate of transits of small evaporating bodies<sup>8–11</sup>, that is, exocomets. Here we report an analysis of more than 1,000 archival spectra gathered between 2003 and 2011, which provides a sample of about 6,000 variable absorption signatures arising from exocomets transiting the disk of the parent star. Statistical analysis of the observed properties of these exocomets allows us to identify two populations with different physical properties. One family consists of exocomets producing shallow absorption lines, which can be attributed to old exhausted (that is, strongly depleted in volatiles) comets trapped in a mean motion resonance with a massive planet. Another family consists of exocomets producing deep absorption lines, which may be related to the recent fragmentation of one or a few parent bodies. Our results show that the evaporating bodies observed for decades in the  $\beta$  Pictoris system are analogous to the comets in our own Solar System.

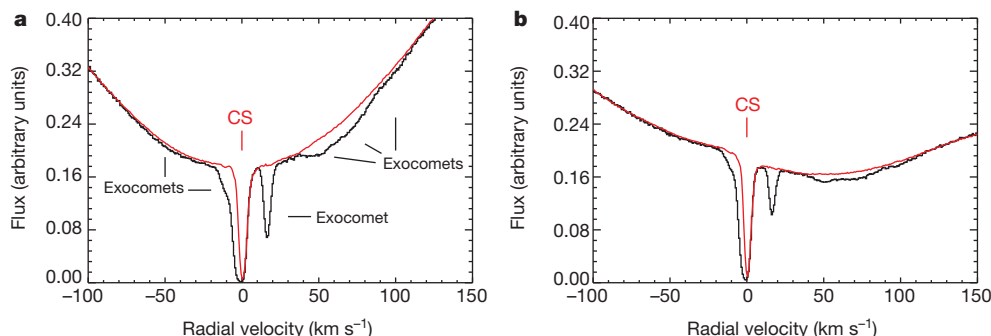
From 2003 to 2011, a total of 1,106 spectra of  $\beta$  Pictoris have been obtained using the HARPS (High Accuracy Radial velocity Planet Searcher) spectrograph. Observations of the calcium (Ca II) doublet—the Ca II K-line at 3,933.66 Å and the Ca II H-line at 3,968.47 Å—show a large number of variable absorption features (Fig. 1) varying on timescales of one to six hours. These features, simultaneously detected in both Ca II K and Ca II H lines, are interpreted as exocomets transiting in front of the stellar disk<sup>7–11</sup>. Since the  $\beta$  Pic Ca II spectrum is typically observed to be stable on 30-min timescales, we averaged together spectra in distinct 10-min time intervals to limit any possible spectral variability. This results in a total of 357 spectra with signal-to-noise ratio greater than 80. To characterize the profile of these transient absorption lines, we divided each of the 357 averaged spectra by a reference spectrum of  $\beta$  Pictoris (Extended Data Figs 1 and 2) assumed to be free of the absorption signatures of transiting exocomets.

Given the HARPS resolution and sensitivity, each  $\beta$  Pictoris spectrum shows an average of about six variable absorption features that are due to exocomets. These features have radial velocities ranging from  $-150$  to  $+200$  km s<sup>-1</sup> with respect to the  $\beta$  Pictoris heliocentric radial velocity ( $\sim 20$  km s<sup>-1</sup>). We fitted each feature with a Gaussian profile and obtained the estimates of  $p_K$  and  $p_H$  (their depths in the Ca II K and H lines),  $v_r$  (the radial velocity of the absorbing cloud) and  $\Delta v$  (the line's full width at half maximum (FWHM) expressed in units of radial velocity).

A large number of the cometary gaseous clouds that pass in front of the star and produce the absorption features observed are smaller than the stellar disk. Therefore, the  $p_K$  and  $p_H$  of each feature depends on  $A$ , the Ca<sup>+</sup> cloud's opacity (absorption depth), and  $\alpha = \Sigma_c / \Sigma_*$ , the ratio of the area of the cloud  $\Sigma_c$  over the area of the stellar disk  $\Sigma_*$ . The simultaneous fit of the K and H lines yields a non-degenerate determination of both  $\alpha$ ,  $A$ ,  $v_r$  and  $\Delta v$  for each feature (Extended Data Fig. 3).

Because the transit of an orbiting exocomet can last several hours, we considered the measurements derived from the fit of only one spectrum per observation day. This ensured that each set of measurements was linked to a different independent object. We thus collected a total of 570 individual sets of measurements from independent transiting cometary clouds. Of these 570 detections, we discarded variable absorption features compatible with  $p_{K,H} < 0.01$  and  $\Delta v < 3$  km s<sup>-1</sup>, in order to avoid contamination introduced by fitting spurious features. We thus end up with a total of 493 detected cometary clouds. For the statistical analysis, we also excluded detections with  $\alpha = 1$ , corresponding to clouds covering the full stellar disk and for which some physical characteristics, like the transit distance, cannot be derived.

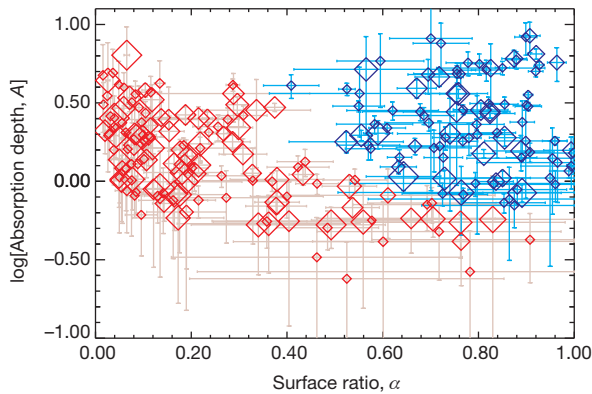
The plot of the absorption depth  $A$  as a function of the surface ratio  $\alpha$  (Fig. 2) shows a depletion of cometary clouds with  $0.2 < \alpha < 0.5$  and  $A \geq 3$  (or  $\log A \geq 0.5$ ). This depletion divides the data into two well separated clusters, revealing the existence of two distinct populations of exocomets. Statistical cluster analysis<sup>12</sup> in the ( $p_H, p_K$ ) diagram (Extended Data Fig. 4) shows that these two populations can be distinguished by the value of  $p_K$ . The first population, called 'population S', corresponding



**Figure 1 | A typical Ca II spectrum of  $\beta$  Pictoris.** **a**, Ca II K-line (3,934 Å). **b**, Ca II H-line (3,968 Å). A typical Ca II spectrum of  $\beta$  Pic (black line) collected on 27 October 2009 is shown together with the derived  $\beta$  Pic stellar spectrum (red line) used as the reference spectrum free of variable absorption features.

Radial velocities are given with respect to the star's rest frame. CS indicates the circumstellar disk contribution, while solid black lines indicate the changes in flux caused by the transiting exocomets. Each transiting exocomet produces an absorption signature detected at the same radial velocity in both Ca II lines.

<sup>1</sup>Centre National de la Recherche Scientifique (CNRS), Unité Mixte de Recherche (UMR) 7095, Institut d'Astrophysique de Paris, 98 bis boulevard Arago, F-75014 Paris, France. <sup>2</sup>Université Pierre et Marie Curie, UMR 7095, Institut d'Astrophysique de Paris, 98 bis boulevard Arago, F-75014 Paris, France. <sup>3</sup>School of Physics and Astronomy, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel. <sup>4</sup>Institut de Radioastronomie Millimétrique, 300 rue de la Piscine, 38406 Saint Martin d'Hères, France. <sup>5</sup>Université Joseph Fourier-Grenoble 1/CNRS—Institut National des Sciences de l'Univers, Institut de Planétologie et d'Astrophysique de Grenoble, UMR 5274, 414 Rue de la Piscine, 38400 St-Martin d'Hères, France.



**Figure 2 | Coma absorption depths as a function of surface ratio for transiting exocomets.** The absorption depths  $A$  (a dimensionless number) of 252 exocomets detected with  $\alpha < 1$  are shown on a logarithmic scale as a function of the surface ratio  $\alpha$ , representing the cloud sizes in units of the stellar disk's area. Small symbols correspond to data taken in 2003 and large symbols to 2011 data. Error bars represent the standard deviation. The 147 exocomets producing shallow absorption lines ( $p_K < 0.4$ ), the so-called population S, are plotted in red, while the 105 exocomets producing deep absorption lines ( $p_K > 0.4$ ), the so-called population D, are plotted in blue. The cloud sizes show a bimodal distribution with a deficiency of exocomets with high absorption depths at intermediate sizes.

mainly to clouds with small surface ratio ( $\alpha \approx 0.1$ ), produces shallow absorption lines ( $p_K < 0.4$ ) and the second population, corresponding to clouds with large surface ratio ( $\alpha \approx 0.8$ ), called 'population D', produces deep absorption lines ( $p_K > 0.4$ ). These two populations present highly different physical properties (Fig. 3).

They have different radial velocity and FWHM distributions<sup>13</sup>. Exocomets of population S have a broad distribution of radial velocities with  $v_{r,S} \approx 36 \pm 55 \text{ km s}^{-1}$ , whereas the population D exocomets have a narrow distribution with  $v_{r,D} \approx 15 \pm 6 \text{ km s}^{-1}$ . Moreover, population S has a broad distribution of FWHM with  $\Delta v_S \approx 55 \pm 55 \text{ km s}^{-1}$ , while population D has a peaked distribution with  $\Delta v_D \approx 7 \pm 3 \text{ km s}^{-1}$ . Since the width of the absorption line is expected to decrease as the distance between the exocomet and host star increases<sup>14</sup>, the bimodal FWHM distribution indicates that population S exocomets transit at shorter

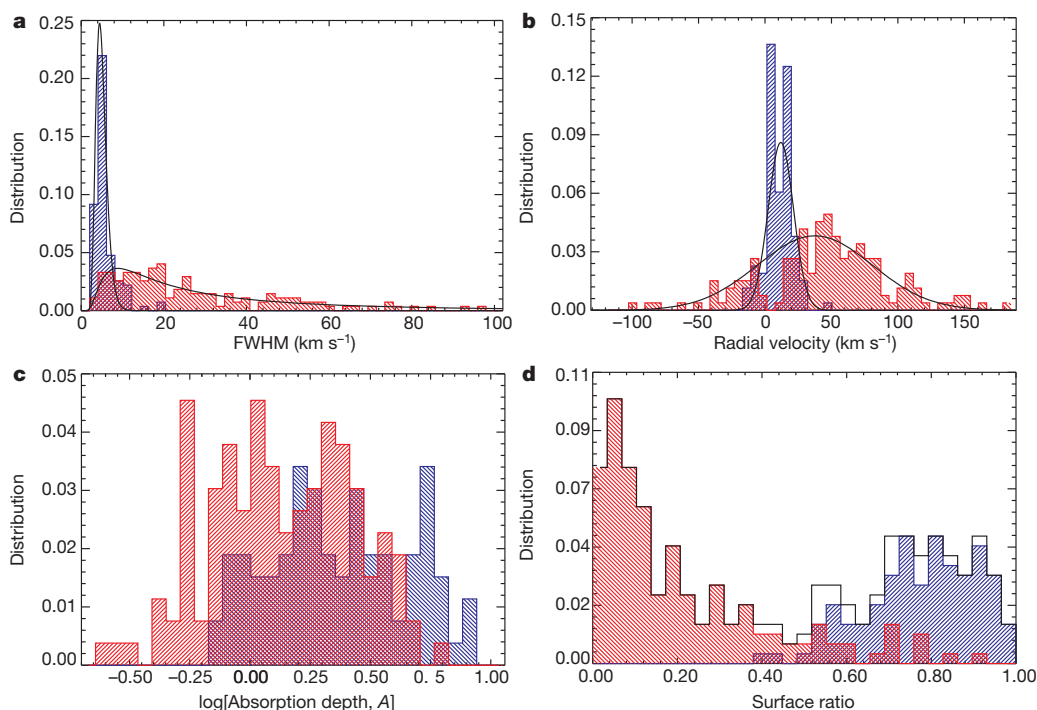
distances than the exocomets of population D. Furthermore, the narrow distribution of radial velocities for population D suggests that these exocomets are gathered on neighbouring orbits with similar longitude of periastron relative to the line of sight<sup>9</sup>, in contrast with the exocomets of population S which seem to be scattered on a wide variety of orbits.

But it is not only different orbital characteristics that distinguish the two populations. The spatial extent of the cometary gaseous cloud and the calcium production rate of a comet both depend on the distance to the star<sup>14,15</sup>, so the resulting Ca II absorption depth also depends on the distance. Thus, if the exocomets of population D had the same intrinsic properties as exocomets of population S and were only transiting at farther distances, then one would expect to measure significantly smaller Ca II absorption depths, as is, however, not observed in Fig. 3c. Moreover, if the observed exocomets were originating from a single family spread over a wide range of orbital distances, we would expect a continuum of measurements in any of the quantities presented in Fig. 3. On the contrary, all histograms show a dichotomy in these measurements, which confirms the existence of two families of exocomets orbiting  $\beta$  Pictoris.

The efficiency of converting stellar irradiance incident on a comet into the evaporation of gas (mainly water) and dust from its core depends on the surface properties and size of its nucleus. In support of the above interpretations, estimates of this evaporation efficiency  $\eta$  show that population D exocomets would have a dust production rate about ten times higher if they were located at the same distance to the star as population S exocomets (Extended Data Fig. 5).

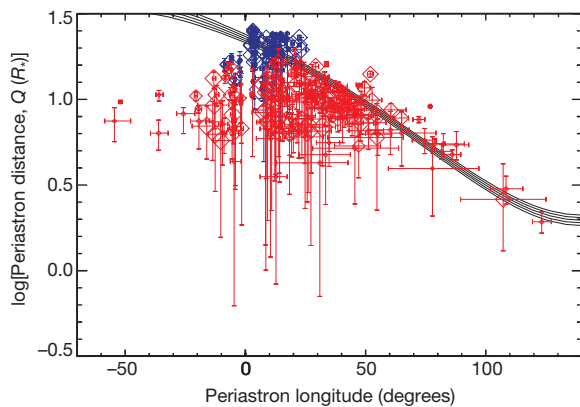
Coupling the evaporation efficiency with a dynamical model of evaporation<sup>14–16</sup> allows the distance to the star at the time of transit  $d$  and the dust production rate  $\dot{M}$  of each detected exocomet to be derived. We find that population D comets orbit twice as far away from  $\beta$  Pic as the population S comets, with  $d_D \approx 19 \pm 4 R_*$  and  $d_S \approx 10 \pm 3 R_*$ , as expected from their lower FWHM (Extended Data Fig. 6), whereas the dust production rates follow  $\dot{M}_D / \dot{M}_S \approx 2$ . These results show that exocomets of population D present more active surfaces than exocomets of population S; this could be explained either by the nuclei being larger in size or by the nuclei being disrupted, thus exposing fresh layers of ice buried in their core.

Given that the radial velocities are directly measured from the spectra, and assuming that exocomets have near-parabolic orbits, the



**Figure 3 | Measured distribution of the physical properties of transiting exocomets.** Histograms showing the distribution of FWHM (a); radial velocity (b); absorption depth (c); and surface ratio  $\alpha$  (d) for the 252 exocomets detected with  $\alpha < 1$ . The 147 exocomets of population S (shown in red) spread over a wide range of FWHM and radial velocities, while the 105 exocomets of population D (shown in blue) produce narrow lines redshifted at low radial velocities. The S population is characterized by small cloud sizes, whereas the D population is composed of large clouds. The cloud sizes for the whole sample of observed exocomets (black line) shows a clear bimodal distribution with a depletion of exocomets at intermediate sizes (d).





**Figure 4 | Periastron distance versus periastron longitude.** The exocomets of population D (shown in blue) have large periastron distances,  $Q_D \approx 18 \pm 4 R_*$ , and a narrow range of longitudes,  $\varpi_D \approx 7 \pm 8^\circ$ . This population of exocomets could originate from the break-up of some parent bodies, which would liberate large amount of fresh volatiles buried in the cometary core, thus doping the gas and dust evaporation rates. The exocomets of population S (shown in red) have smaller periastron distances,  $Q_S \approx 9 \pm 3 R_*$ , and a wide range of longitudes,  $\varpi_S \approx 22 \pm 25^\circ$ . The  $Q$ -versus- $\varpi$  relationship predicted for comets trapped in 4:1 mean motion resonance with a massive planet<sup>18</sup> ( $m' \approx 10 M_J$ ,  $a' \approx 4.5$  AU,  $\varpi' \approx -40^\circ$ ,  $e' \approx 0.04$ ) is shown as black solid lines. Error bars represent the standard deviation.

combination of the distance and the radial velocity yields  $Q$ , the periastron distance, and  $\varpi$ , the longitude of the periastron. The plot of  $Q$  as a function of  $\varpi$  shows that orbital properties also differ between the two families (Fig. 4). Exocomets of population D have larger periastron distances than exocomets of population S, with  $Q_D \approx 18 \pm 4 R_*$  and  $Q_S \approx 9 \pm 3 R_*$ . They also present a narrow distribution of longitude of periastron, indicating that all population D exocomets share similar orbits, with  $\varpi_D \approx 7 \pm 8^\circ$ . This concentration of a large number of bodies on similar orbits with a nearly constant longitude of periastron can be explained by the disruption of one or a few individual exocomets. These observations resemble the Kreutz Family comets in our own Solar System<sup>17</sup>, which are detected with periastron distances ranging from 0.005 AU to 0.01 AU and periastron longitude ranging from  $10^\circ$  to  $90^\circ$ .

Conversely, the population S follows, as expected, a much broader distribution of longitude of periastron, with  $\varpi_S \approx 22 \pm 25^\circ$ . The arc-like structure in Fig. 4 suggests that a fraction of the exocomets of population S present a strong correlation between the periastron distances and the longitudes: starting at  $\log Q \approx 1.0$  and  $\varpi \approx 50^\circ$ , the periastron distances decrease to  $\log Q \approx 0.4$  and  $\varpi \approx 100^\circ$ . This is exactly the behaviour predicted<sup>18</sup> for comets trapped in a mean motion resonance with a massive planet (see figure 11 of ref. 18) such as  $\beta$  Pic b (refs 19 and 20). In this scenario, the lower evaporation rate of population S exocomets is explained by the exhaustion of volatiles at the surface of their nucleus caused by a large number of periastron passages as they evolve towards high eccentric orbits<sup>21,22</sup>.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 10 March; accepted 8 September 2014.**

1. Roberge, A. *et al.* High-resolution Hubble Space Telescope STIS spectra of C I and CO in the  $\beta$  Pictoris circumstellar disk. *Astrophys. J.* **538**, 904–910 (2000).

2. Lecavelier des Etangs, A. *et al.* Deficiency of molecular hydrogen in the disk of  $\beta$  Pictoris. *Nature* **412**, 706–708 (2001).
3. Okamoto, Y. K. *et al.* An early extrasolar planetary system revealed by planetesimal belts in  $\beta$  Pictoris. *Nature* **431**, 660–663 (2004).
4. Telesco, C. M. *et al.* Mid-infrared images of  $\beta$  Pictoris and the possible role of planetesimal collisions in the central disk. *Nature* **433**, 133–136 (2005).
5. Roberge, A., Feldman, P. D., Weinberger, A. J., Deleuil, M. & Bouret, J.-C. Stabilization of the disk around  $\beta$  Pictoris by extremely carbon-rich gas. *Nature* **441**, 724–726 (2006).
6. de Vries, B. L. *et al.* Comet-like mineralogy of olivine crystals in an extrasolar proto-Kuiper belt. *Nature* **490**, 74–76 (2012).
7. Vidal-Madjar, A., Lecavelier des Etangs, A. & Ferlet, R.  $\beta$  Pictoris, a young planetary system? A review. *Planet. Space Sci.* **46**, 629–648 (1998).
8. Ferlet, R., Vidal-Madjar, A. & Hobbs, L. M. The Beta Pictoris circumstellar disk. V—Time variations of the CA II-K line. *Astron. Astrophys.* **185**, 267–270 (1987).
9. Beust, H., Vidal-Madjar, A., Ferlet, R. & Lagrange-Henri, A. M. The Beta Pictoris circumstellar disk. X—Numerical simulations of infalling evaporating bodies. *Astron. Astrophys.* **236**, 202–216 (1990).
10. Beust, H., Vidal-Madjar, A., Ferlet, R. & Lagrange-Henri, A. M. The Beta Pictoris circumstellar disk. XI—New CA II absorption features reproduced numerically. *Astron. Astrophys.* **241**, 488–492 (1991).
11. Vidal-Madjar, A. *et al.* HST-GHRS observations of  $\beta$  Pictoris: additional evidence for infalling comets. *Astron. Astrophys.* **290**, 245–258 (1994).
12. Everitt, B. S. *Cluster Analysis* (Halsted Press, 1993).
13. Lagrange, A.-M. *et al.* The  $\beta$  Pictoris circumstellar disk. XXI. Results from the December 1992 spectroscopic campaign. *Astron. Astrophys.* **310**, 547–563 (1996).
14. Beust, H., Lagrange, A.-M., Plazy, F. & Mouillet, D. The  $\beta$  Pictoris circumstellar disk. XXII. Investigating the model of multiple cometary infalls. *Astron. Astrophys.* **310**, 181–198 (1996).
15. Beust, H. & Tagger, M. A hydrodynamical model for infalling evaporating bodies in the Beta Pictoris circumstellar disk. *Icarus* **106**, 42–58 (1993).
16. Huebner, W. F. Physics and chemistry of comets. In *Physics and Chemistry of Comets* Ch. 2 (Springer, 1990).
17. Marsden, B. G. The sungrazing comet group. *Astron. J.* **72**, 1170 (1967).
18. Beust, H. & Morbidelli, A. Mean-motion resonances as a source for infalling comets toward  $\beta$  Pictoris. *Icarus* **120**, 358–370 (1996).
19. Lagrange, A.-M. *et al.* A probable giant planet imaged in the  $\beta$  Pictoris disk. VLT/NaCo deep L'-band imaging. *Astron. Astrophys.* **493**, L21–L25 (2009).
20. Lecavelier des Etangs, A. & Vidal-Madjar, A. Is  $\beta$  Pictoris b the transiting planet of November 1981? *Astron. Astrophys.* **497**, 557–562 (2009).
21. Thébaud, P. & Beust, H. Falling evaporating bodies in the  $\beta$  Pictoris system. Resonance refilling and long term duration of the phenomenon. *Astron. Astrophys.* **376**, 621–640 (2001).
22. Beust, H. & Morbidelli, A. Falling evaporating bodies as a clue to outline the structure of the  $\beta$  Pictoris young planetary system. *Icarus* **143**, 170–188 (2000).

**Acknowledgements** This work was based on observations made with ESO telescopes at the La Silla Observatory and data obtained from the ESO Science Archive Facility. This work has been supported by an award from the Fondation Simone et Cino Del Duca. We also acknowledge support from the French Agence Nationale de la Recherche (ANR), under programmes ANR-12-BS05-0012 Exo-Atmos and ANR-2010 BLAN-0505-01 (EXOZODI). We thank P. A. Wilson for his comments on the manuscript.

**Author Contributions** F.K. led the data analysis, with contributions from A.L.d.E., J.B., A.V.-M., R.F. and G.H.; F.K. and A.L.d.E. wrote the paper; H.B. computed mean-motion resonance curves and developed theoretical modelling; A.L.d.E., A.V.-M., A.-M.L. and H.B. contributed to the conception of the project. All authors discussed the results and commented on the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to F.K. ([kiefer@iap.fr](mailto:kiefer@iap.fr)) or A.L.d.E. ([lecavel@iap.fr](mailto:lecavel@iap.fr)). The data used for this study come from the ESO database (<http://archive.eso.org>) and are referenced with the following IDs: 060.A-9036(A), 072.C-0636(A), 072.C-0636(B), 075.C-0234(A), 075.C-0234(B), 076.C-0073(A), 076.C-0279(A), 078.C-0209(A), 078.C-0209(B), 080.C-0032(A), 080.C-0032(B), 080.C-0664(A), 080.C-0712(A), 081.C-0034(B), 082.C-0308(A), 082.C-0412(A), 084.C-1039(A), 184.C-0815(C) and 184.C-0815(F).

## METHODS

This section describes the data analysis, the derivation of the physical properties of the exocomets and details of the interpretation.

**Spectra.** HARPS is a high-resolution (about 115,000), high-precision ( $< 1 \text{ m s}^{-1}$ ) spectrograph installed at the 3.6-m European Southern Observatory (ESO) telescope located at La Silla, Chile. The spectra were extracted from the detector images with the Data Reduction System pipeline of HARPS, which includes localization of the spectral orders on the two-dimensional images, optimal order extraction, cosmic-ray rejection, wavelength calibration, flat-field corrections, and one-dimensional reconnection of the spectral orders after correction for the blaze. A typical HARPS spectrum of  $\beta$  Pictoris includes the Ca II doublet lines (3,933.66 Å and 3,968.47 Å), which show contributions from (1) the rotationally broadened stellar lines, (2) the circumstellar gaseous disk, and (3) variable absorption features (Fig. 1).

To compare  $\beta$  Pictoris spectra collected at different epochs, we first normalized all the spectra to the same mean flux level using the mean of the flux in the wings of the Ca II line, where no variable absorption features are present. We then checked for possible shifts in wavelength calibration with time by using the Na I lines as a reference. The circumstellar Na I line is steep (Extended Data Fig. 1) and confirms the tremendous stability of the instrument during the observation campaign—as expected for this spectrograph, which is aimed at detecting minute radial velocity variations. Table 2 in ref. 23 shows that the accuracy of HARPS is better than  $1 \text{ m s}^{-1}$  over several years. We thus have high confidence in the detected spectral variations.

We also checked for stellar variations in the Ca II spectrum over long timescales. To do so, we computed a reference spectrum of the Ca II stellar lines (see next paragraph) for each observational campaign (2003–2004, 2004–2008, 2008–2009, 2009–2011 and 2011). A thorough comparison of these reference spectra allowed us to exclude any significant variation of the stellar lines between 2003 and 2011. We hence decided to use the whole set of spectra to compute one common reference spectrum.

However, variations of the circumstellar linewidth by about  $3 \text{ km s}^{-1}$  are seen between 2003 and 2011. Since the HARPS spectrograph is not capable of resolving features below  $3 \text{ km s}^{-1}$  at the Ca II doublet's wavelengths, we discarded the part of the spectra corresponding to the circumstellar line region extending from  $18 \text{ km s}^{-1}$  to  $24 \text{ km s}^{-1}$  around the circumstellar line centre at  $21 \text{ km s}^{-1}$ , the systemic radial velocity of  $\beta$  Pic.

**Derivation of the reference spectrum.** To characterize the absorption features, we divided each observed  $\beta$  Pic spectrum  $F_{\text{obs}}(\lambda)$  by a reference spectrum  $F_{\text{ref}}(\lambda)$ . The reference spectrum was obtained as described below and includes both stellar and circumstellar absorption components in the Ca II doublet. In the absence of transiting exocomets, the normalized spectrum  $F_{\text{obs}}/F_{\text{ref}}$  is constant and equal to 1.

Each Ca II spectrum of  $\beta$  Pic shows many variable absorption features of different depth and width. As a consequence, none of the 1,106 spectra could be considered as an estimate of the reference spectrum. However, at any given wavelength  $\lambda_i$  around the stellar Ca II K and H lines, variable features appear and disappear randomly. As a result, amongst the whole set of flux measurements  $F_k(\lambda_i)$  (for  $k = 1, \dots, N = 1,106$ ), a small fraction has no or little contamination by variable features, and can be used to compute the reference spectrum. We further assumed that, in the absence of variable absorption features, the noise  $\delta F$  in the flux measurements is Gaussian, and we checked that the root mean square (r.m.s.) of the measurements is proportional to  $\sqrt{F}$ , with a constant factor independent of the wavelength.

First, we obtained an estimation of the reference spectrum by considering at each wavelength  $\lambda_i$  only the 2.5% highest flux values. Each measurement was subsequently sorted in a decreasing order such that  $F_k(\lambda_i) \geq F_{k+1}(\lambda_i)$  (for  $k = 1, \dots, N = 1,105$ ). The 2.5% highest flux values ( $k \leq 28$ ) are probably not contaminated by the variable absorption features and can be considered as an upper limit randomly drawn from the Gaussian distribution of the noise centred around the true reference flux. In this case, we can estimate the difference between any  $F_k(\lambda_i)$  and the reference spectrum  $F_{\text{ref}}(\lambda_i)$ . Above  $F_k(\lambda_i)$  there are  $k$  flux measurements that are a fraction  $k/N$  of the total number of flux measurements at the given wavelength. Assuming that the noise  $\delta F$  is Gaussian with first momentum  $\sigma = \text{r.m.s.}$ , we compute the cut-off value  $\alpha_k$  at which the probability that  $\delta F > \alpha_k \times \text{r.m.s.}$  is  $k/N$ . We then obtain a first estimate of the reference flux:

$$F_{\text{ref},k}^{(1)}(\lambda_i) = F_k(\lambda_i) - \alpha_k \times \text{r.m.s.} \quad (1)$$

However, variable absorption features appear randomly at each wavelength, disturbing the derivation of the reference spectrum. To improve it, we introduced a second step in the computation using a larger number of flux values. For each  $k$  and wavelength  $\lambda_i$ , we computed the mean value  $\bar{F}_k(\lambda_i)$  of the flux measurements  $F_p(\lambda_i)$  for  $p$  such that  $F_{\text{ref},k}^{(1)}(\lambda_i) < F_p(\lambda_i) < F_k(\lambda_i)$ . These flux measurements exceed the first estimate of the reference spectrum and are therefore not likely to be affected by variable absorption lines. This new step takes into account up to 500 flux measurements in the computation of the reference flux, to be compared to the 28 most

extreme flux values used in the first step. Using the value  $\bar{F}_k(\lambda_i)$ , we obtain a new estimate of the reference spectrum given by:

$$F_{\text{ref},k}^{(2)} = \bar{F}_k(\lambda_i) - \beta_k \times \text{r.m.s.} \quad (2)$$

where  $\beta_k$  is the average value of a normalized Gaussian variable in the range  $[0, \alpha_k]$ , given by:

$$\beta_k = \frac{\int_0^{\alpha_k} x e^{-x^2/2} dx}{\int_0^{\alpha_k} e^{-x^2/2} dx} \quad (3)$$

Finally, we compute the final reference spectrum by taking the average of all the  $F_{\text{ref},k}^{(2)}$  with  $k$  varying from 3 to 28:

$$F_{\text{ref}} = \left\langle F_{\text{ref},k}^{(2)} \right\rangle_{3 \leq k \leq 28} \quad (4)$$

The application of this three-step method at all wavelengths  $\lambda_i$  of the Ca II spectrum allowed us to obtain the reference spectrum plotted in Fig. 2. Its accuracy is such that a small interstellar line is detected on the left of the circumstellar line<sup>24</sup>. We performed at each wavelength a  $\chi^2$  test to compare the tail distribution of flux measurements going from  $F_{\text{ref}}$  to  $F_{\text{ref}} + 3\sigma$  with a Gaussian distribution of photon noise. The agreement is good, with 87% of the reference flux values passing the test at a significance level of 5%.

**Fitting method.** We obtained normalized spectra by dividing each spectrum by the reference spectrum. The normalized spectra show exclusively variable absorption features, as can be seen by comparing Extended Data Figs 1 and 3. Spectroscopic variations were typically not seen on timescales less than 30 min (corresponding to the minimum duration of a transit); however, to limit the effects of any possible spectral variability, the spectra were averaged into separate 10-min time intervals. We thus obtained 357 re-sampled spectra with signal-to-noise ratio  $> 80$ . Each of them contains an average of six variable absorption features with radial velocities between  $-150 \text{ km s}^{-1}$  and  $200 \text{ km s}^{-1}$  with respect to  $\beta$  Pic. These features can be fitted by a Gaussian profile:

$$p_{K,H} e^{-4 \ln 2 \frac{(v-v_0)^2}{\Delta v^2}} \quad (5)$$

where  $p_K$  and  $p_H$  are the line depths in the Ca II K and H lines, respectively,  $v_0$  is the radial velocity of the coma and  $\Delta v$  is the line's FWHM. Because the absorption features are produced by gaseous clouds passing in front of the star, the depths  $p_K$  and  $p_H$  are related to the cloud-to-star surface ratio  $\alpha = \Sigma_c / \Sigma_* \leq 1$  and the optical depth at the centre of the absorption feature in the Ca II K line, or absorption depth  $A$ , by:

$$\begin{aligned} p_K &= \alpha(1 - e^{-A}) \\ p_H &= \alpha(1 - e^{-A/2}) \end{aligned} \quad (6)$$

The absorption depth  $A$  depends on the density and depth of the medium. This quantity is directly related to the  $p_K/p_H$  ratio characterizing the saturation in absorption within the cloud:

$$\frac{p_K}{p_H} = 1 + e^{-\frac{A}{2}} \quad (7)$$

With  $\alpha \leq 1$ , the relationship between equations (6) and (7) becomes:

$$1 \leq \frac{p_K}{p_H} \leq 2 - p_H \quad (8)$$

We fitted each variable absorption feature simultaneously in the Ca II K line and Ca II H line of all 357 normalized spectra, providing non-degenerate estimates of  $\alpha$ ,  $A$ ,  $v_0$  and  $\Delta v$  for each variable absorption component. The huge number of spectra and the large number of blended components in each spectrum makes the fitting of these features challenging. We developed a systematic procedure which we used to fit each spectrum automatically by searching for as many lines as required to best-fit the data. Since the prior on the number of components is a uniform probability function, we used the Bayesian information criterion (BIC) to get the optimal number of components necessary to build a fair model of the normalized spectrum. The BIC is defined as:  $\text{BIC} = \chi^2 + k \ln N$ , where  $N$  is the total number of data points and  $k$  is the number of parameters. For each spectrum, we took a number  $n$  of components yielding a BIC with a value  $\text{BIC}_n$  such that the fit with an additional component yields a  $\text{BIC}_{n+1}$  that does not decrease by more than 6. When  $\text{BIC}_n$  minus  $\text{BIC}_{n+1}$  is less than 6, the model with  $n+1$  components has a less than 95% probability of being closer to reality than model with  $n$  components. A typical example of a resulting fit using this procedure is plotted in Extended Data Fig. 3.

**Separation of the populations.** Figure 2 reveals the presence of two clusters of data separated at  $\log A \approx 0.5$ ; one at small surface ratio ( $\alpha < 0.2$ ) and the other one at a larger surface ratio ( $\alpha > 0.5$ ). Extended Data Fig. 4 also suggests the presence of two clusters of features in the distribution of the depth of the K and H lines. We

performed a statistical cluster analysis using the  $k$ -mean cluster algorithm<sup>12</sup> in the  $(p_H, p_K)$  diagram. This algorithm identified two clusters of data with  $p_K < 0.4$  on one side and  $p_K > 0.4$  on the other side. These two clusters are directly related to the two clusters in Fig. 2. We performed a Kolmogorov–Smirnov test to compare the distributions in parameters  $\alpha$ ,  $A$ ,  $v_i$  and  $\Delta v$  of these two clusters. For each of these parameters, the two distributions are different with a probability  $P > 0.9999$ . A similar statistical cluster analysis performed in  $(A, \alpha)$  space led to an analogous separation of the two populations.

**Evaporation efficiency.** The quantities measured in the ionic  $\text{Ca}^+$  cloud transiting in front of  $\beta$  Pictoris can be used to derive the physical properties of the exocomets such as their distance to the star and the dust production rate in their coma. To do so, we estimate the evaporation efficiency of each individual detected exocomet, a quantity describing the efficiency with which a comet captures and reprocesses input stellar energy flux into dust and gas evaporation from its nucleus.

*Definition.* After being damped by the opacity of the dust coma, the stellar radiation incident on the comet reaches the icy nucleus surface covered by a thin layer of agglomerated dust<sup>16</sup>, and heats it up. A significant part of this heat energy is used to sublimate water from the ice of the nucleus, at a rate  $Z_{\text{H}_2\text{O}}$ ; the remaining energy is absorbed by the crust or re-emitted by the surface.

The heat re-emitted or absorbed by the dusty surface is in part used to thermalize the sublimated gas. Assuming that the distance  $d$  of a comet to  $\beta$  Pic is around  $10R_*$  (where  $R_*$  is the stellar radius) and given that  $T_{\text{eff}} \approx 8,000$  K, the temperature at the surface of the comet is  $T_c = T_{\text{eff}} \sqrt{R_*/d}$ . As the gas heats up, it escapes from the nucleus surface and flows out radially. A water molecule heated to 2,500 K is typically accelerated to a radial velocity of  $1 \text{ km s}^{-1}$  provided that  $v_i \approx \sqrt{3k_B T/m_{\text{H}_2\text{O}}}$ , where  $k_B$  is the Boltzmann constant,  $T$  is the temperature and  $m_{\text{H}_2\text{O}}$  is the mass of one water molecule. While escaping, the gas picks up dust grains from the dust mantle and drags them outward, with a mass rate  $\dot{M}$ . Kinetic energy is thus transferred from the gas to the dust grains, which are then accelerated to the expansion velocity of water molecules near the nucleus<sup>14,25</sup>,  $v_n \approx 1 \text{ km s}^{-1}$ .

The efficiency of this evaporation process, which depends on the surface properties of the nucleus (such as size, albedo and fragmentation), can be measured by comparing the total energy used per unit time to evaporate dust at a mass rate  $\dot{M}$  with the input energy per unit of time reaching the comet, that is, the stellar flux. This leads to the following definition of the evaporation efficiency;

$$\eta = \log \frac{Z_{\text{H}_2\text{O}} L_{\text{H}_2\text{O}} + \frac{1}{2} \dot{M} v_n^2}{F_*(d)} \quad (9)$$

We introduced  $L_{\text{H}_2\text{O}} \approx 3 \times 10^2 \text{ kJ kg}^{-1}$  as the latent heat of vaporization for water, and  $F_*(d)$  is the stellar flux at a distance  $d$ , which is related to the stellar luminosity  $L_*$  by  $F_* = L_*/4\pi d^2$ , with  $L_* = 8.6 L_\odot$ . We neglected the gas kinetic energy, which is an order of magnitude lower than the latent heat of water sublimation.

Reducing numerical factors, and assuming the dust-to-gas mass ratio is constant and close to 1 such that  $Z_{\text{H}_2\text{O}} \approx \dot{M}$ , the evaporation efficiency can be expressed as:

$$\eta \approx \log(\dot{M} d^2) - 1.9 \quad (10)$$

The distance  $d$  is expressed in stellar radius units and  $\dot{M}$  in  $\text{kg s}^{-1}$ . Typical values<sup>9</sup> are expected to be  $10^7 < \dot{M} < 5 \times 10^8 \text{ kg s}^{-1}$  and  $10 < d < 50 R_*$ , yielding  $7 < \eta < 10$ . *Measurements of the evaporation efficiency.* The evaporation efficiency can be estimated from the measured values of  $\alpha$  and  $A$  using the conservation of momentum in the cometary cloud: the total momentum carried by the stellar photons which are absorbed by the  $\text{Ca}^+$  cloud equates the total momentum gained by these  $\text{Ca}^+$  atoms.

On one side, we thus consider the amount of photons absorbed by the  $\text{Ca}^+$  cloud, accounting for the contribution of both K and H lines, integrated over its projected surface, to be  $\Sigma_c = \pi R_c^2 \alpha$  and  $F_{*,d}$  is the stellar flux at distance  $d$  from the star. This leads to an absorbed momentum flux:

$$\frac{dP}{dt} \Big|_{\text{abs}} = \Sigma_c \int (1 - e^{-A}) e^{-4 \ln 2 \frac{(v-v_0)^2}{\Delta v^2}} F_{*,d}(\lambda) \frac{d\lambda}{c} \quad (11)$$

$$+ \Sigma_c \int (1 - e^{-A/2}) e^{-4 \ln 2 \frac{(v-v_0)^2}{\Delta v^2}} F_{*,d}(\lambda) \frac{d\lambda}{c} \quad (12)$$

with  $v - v_0 = c(\lambda - \lambda_0)/\lambda_0$  and  $\lambda_0 = 3,950 \text{ \AA}$ .

On the other side, we consider the total momentum  $P$  gained per unit of time by the  $\text{Ca}^+$  ions when they are accelerated from their initial velocity  $v_i$  to a final velocity  $v_f$ :

$$\frac{dP}{dt} \Big|_{\text{acc}} = \dot{M}_{\text{Ca}^+} \langle v_f - v_i \rangle_{v_f} \quad (13)$$

We assume  $v_f$  to follow a distribution of typical width  $\Delta v$  and mean  $v_0$ . The variation of momentum is averaged with respect to this distribution, with  $\langle v_f \rangle = v_0$ .

These two independent expressions for the momentum flux can be computed for each line as a function of  $A$ ,  $\alpha$  and  $\Delta v$  following:

$$\begin{aligned} \frac{dP}{dt} \Big|_{\text{abs}} &= \alpha \pi R_*^2 \frac{(19.3 \text{ pc})^2 F_\oplus(\lambda_0) \lambda_0 \Delta v}{d^2 R_*^2 c^2} \sqrt{\frac{\pi}{4 \ln 2}} (2 - e^{-A} - e^{-A/2}) \\ \frac{dP}{dt} \Big|_{\text{acc}} &= \dot{M}_{\text{Ca}^+} (v_0 - v_i) \approx \dot{M}_{\text{Ca}^+} \times \Delta v \end{aligned} \quad (14)$$

where  $F_\oplus(\lambda_0) = 0.25 \times 10^{-10} \text{ erg cm}^{-2} \text{ s}^{-1} \text{ \AA}^{-1}$  is the  $\beta$  Pic stellar flux measured from Earth around wavelength  $\lambda_0 \approx 3,950 \text{ \AA}$ , accounting for a 0.25 reduction factor at the bottom of the K and H lines; the distance  $d$  is expressed in stellar radius units.

Momentum conservation implies that  $dP/dt|_{\text{acc}} = dP/dt|_{\text{abs}}$ , leading to:

$$\dot{M}_{\text{Ca}^+} d^2 \approx 1.3 \times 10^9 \alpha (2 - e^{-A} - e^{-A/2}) \quad (15)$$

Taking into account the expected abundance of calcium in silicate<sup>9</sup> ( $\dot{M}_{\text{Ca}^+} \approx 0.01 \dot{M}_{\text{dust}}$ ), we conclude that:

$$\eta \approx 9.2 + \log[\alpha(2 - e^{-A} - e^{-A/2})] \quad (16)$$

*Discussion.* The distribution of the measured evaporation efficiencies for the observed exocomets is in good agreement with typically expected values,  $7 < \eta < 10$  (Extended Data Fig. 5). The two populations of exocomets have distinct evaporation efficiency distributions with  $\eta_S \approx 8.6 \pm 0.4$  and  $\eta_D \approx 9.4 \pm 0.1$ . Exocomets of population D are thus almost ten times more efficient at capturing and converting input stellar energy into gas and dust evaporation than exocomets of population S. In other words, they would have a dust production rate ten times higher, if they were located at the same distance to the star as population S.

The uncertainties quoted above on the measure of the evaporation efficiency in each population do not include the effects of the model's approximations, in particular in the estimation of the cometary cloud area and the velocity of the  $\text{Ca}^+$  ions. We estimated that model uncertainties by a factor of about two or less on these two quantities lead to an error bar of about  $\pm 0.3$  for the evaporation efficiency, which is about three times smaller than the difference seen between the two populations ( $\Delta \eta \approx 0.8$ ). This is therefore additional evidence of the existence of two families of exocomets with distinct intrinsic properties.

**Distance and dust production rate.** Using the model of ref. 15, we can derive the distance between the star and the exocomets at the time of their transit. In this model, ions are supersonically dragged by the evaporating gas flowing out from the nucleus. At a large distance from the nucleus, the radially expanding ions are slowed down by the radiation pressure to a subsonic velocity and are further accelerated in the anti-stellar direction.

As a result, a shock surface is formed at distance  $r_0$  from the nucleus:

$$r_0 = \left( \frac{\gamma F \dot{M} d^2}{8\pi \beta G M_* a_1} \right)^{1/3} \quad (17)$$

We introduced  $\gamma = 0.01$ , the typical calcium abundance in silicate grains;  $\beta = 77$ , the ratio of radiation pressure to gravity for  $\text{Ca}^+$  ions in the  $\beta$  Pic environment;  $M_* = 1.7 M_\odot$ , the mass of  $\beta$  Pic; and  $a_1 = \mu/\sigma$ , a constant factor accounting for the shock surface<sup>15</sup> and dependent on  $\mu$ , the  $\text{Ca}^+$  mass, and  $\sigma$ , the effective cross-section of the stellar flux absorption by the  $\text{Ca}^+$  ions. The  $F$  factor<sup>15</sup> depends on the gas production rate  $Z$  (typically equal to dust production rate  $\dot{M}$  in Solar System comets):

$$F = \frac{\dot{M}}{2v_e \mu m_{\text{H}_2\text{O}}} \sqrt{\frac{4m_{\text{H}} \alpha_{\text{H}} e^2}{4\pi \epsilon_0}} \quad (18)$$

where  $m_{\text{H}_2\text{O}}$  is the mass of the water molecule,  $m_{\text{H}}$  is the hydrogen mass;  $\alpha_{\text{H}} = 6.67 \times 10^{-31} \text{ m}^3$  is the polarizability of the hydrogen atom; and  $v_e \approx 10 \text{ km s}^{-1}$  is the dust expansion velocity just below the shock surface. The effective cross-section of stellar flux absorption by the  $\text{Ca}^+$  ions is

$$\sigma = \sum_{\text{K,H}} \frac{\lambda_{\text{K,H}}}{\Delta v} \frac{1}{4\pi \epsilon_0} \frac{\pi e^2}{m_e c} f_{\text{K,H}} \quad (19)$$

introducing  $f_K = 0.69$  and  $f_H = 0.34$ , the oscillator strengths of the  $\text{Ca II K}$  and H lines; and  $\Delta v \approx 1 \text{ km s}^{-1}$ , the estimated transition linewidth, taking into account the natural, thermal and collisional broadenings. Provided  $r_0 \leq R_*$ , the surface ratio  $\alpha = \Sigma_c/\Sigma_*$  is roughly given by  $\alpha \approx r_0^2/R_*^2$ , yielding:

$$\alpha = 4.5 \times 10^{-14} \dot{M}^{4/3} d^{4/3} \quad (20)$$



Using equation (15), we obtain the distance of each exocomet to  $\beta$  Pic at the time of the transit given in stellar radius:

$$d = 6.2 \times 10^{-9} 10^{\eta} \alpha^{-3/4} \quad (21)$$

Our measurements lead to estimates of the distance between  $1R_*$  and  $30R_*$  (Extended Data Fig. 6), as expected. The dust production rate  $\dot{M}$  is then deduced from equations (21) and (10).

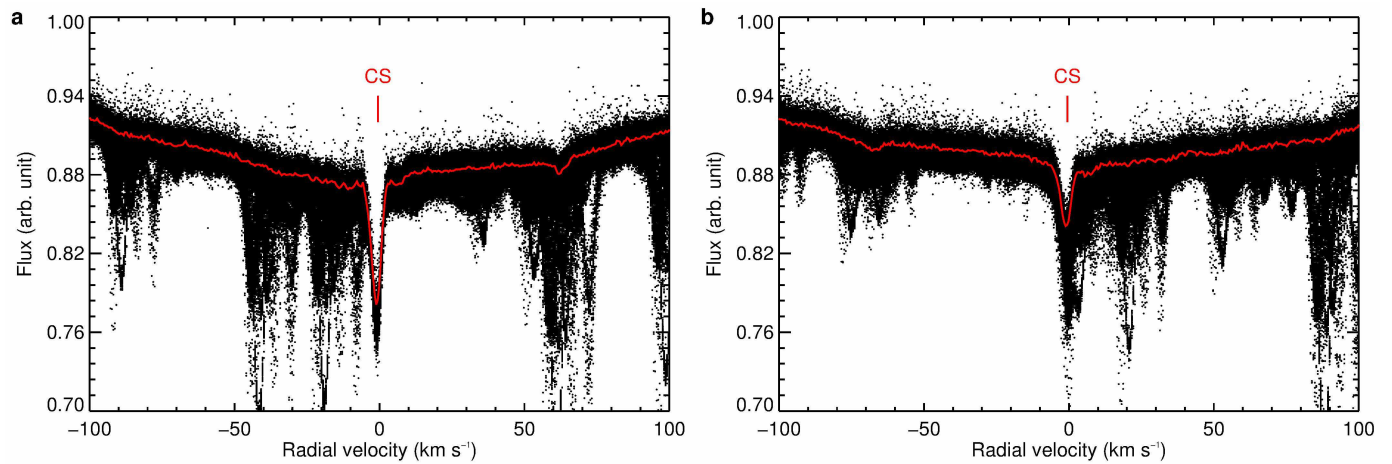
**Periastron distance and longitude.** Assuming that each exocomet exhibits a near-parabolic orbit, an estimate of the distance to the star, together with the measurement of the radial velocity at the time of the transit, allows an estimation of the periastron orientation and distance. We define  $\varpi$  to be the longitude of the periastron, which is the true anomaly of the line of sight with respect to the exocomet's periastron, and  $Q$  to be the periastron distance in units of stellar radius,  $R_*$ . These two quantities can be expressed with respect to the distance and the radial velocity by solving:

$$v_r = \sqrt{\frac{G M_*}{R_* d}} \frac{\sin \varpi}{\sqrt{1 + \cos \varpi}}$$

$$d = \frac{2 Q}{1 + \cos \varpi} \quad (22)$$

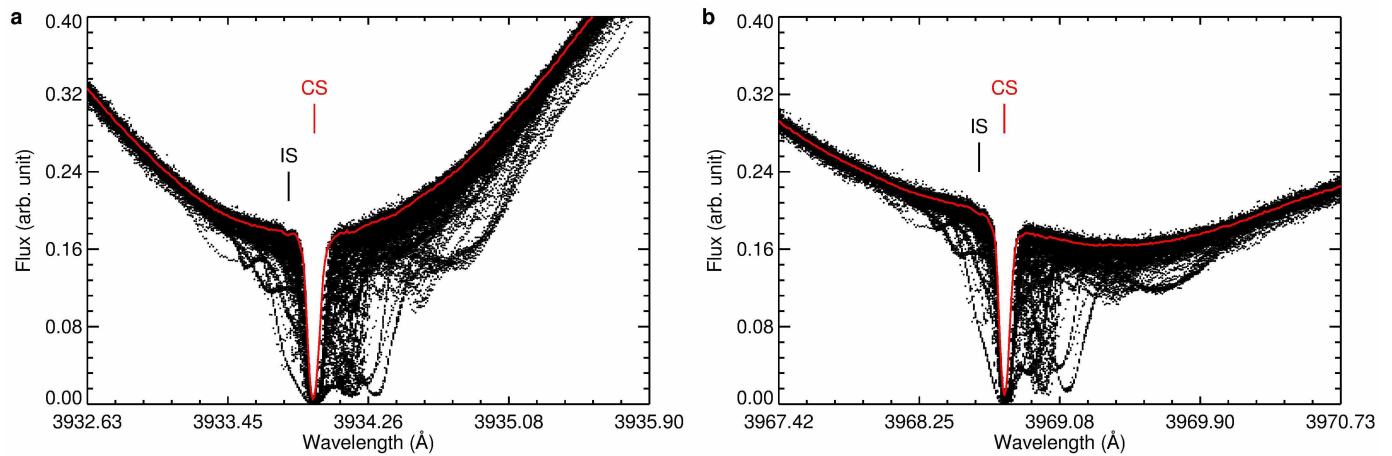
The first equation is solved using a numerical inversion method, and the second is solved once  $\varpi$  is known. We plot the  $(\varpi, \log Q)$  diagram in Fig. 4.

23. Pepe, F. *et al.* The HARPS search for Earth-like planets in the habitable zone. I. Very low-mass planets around HD 20794, HD 85512, and HD 192310. *Astron. Astrophys.* **534**, A58 (2011).
24. Lallement, R., Ferlet, R., Lagrange, A. M., Lemoine, M. & Vidal-Madjar, A. Local cloud structure from HST-GHRS. *Astron. Astrophys.* **304**, 461–474 (1995).
25. Beust, H., Lagrange-Henri, A. M., Vidal-Madjar, A. & Ferlet, R. The Beta Pictoris circumstellar disk. IX—Theoretical results on the infall velocities of CA II, AL III, and MG II. *Astron. Astrophys.* **223**, 304–312 (1989).



**Extended Data Figure 1 | The Na I spectrum of  $\beta$  Pictoris.** **a**, Na I D2-line spectrum (5,890 Å). **b**, Na I D1-line spectrum (5,896 Å). It shows the superposition of all Na I spectra of  $\beta$  Pic (black dots) compared with the stellar reference spectrum (red line). Radial velocities are given in the star's rest frame. The stable Na I line centred at the star's radial velocity is identified as due to

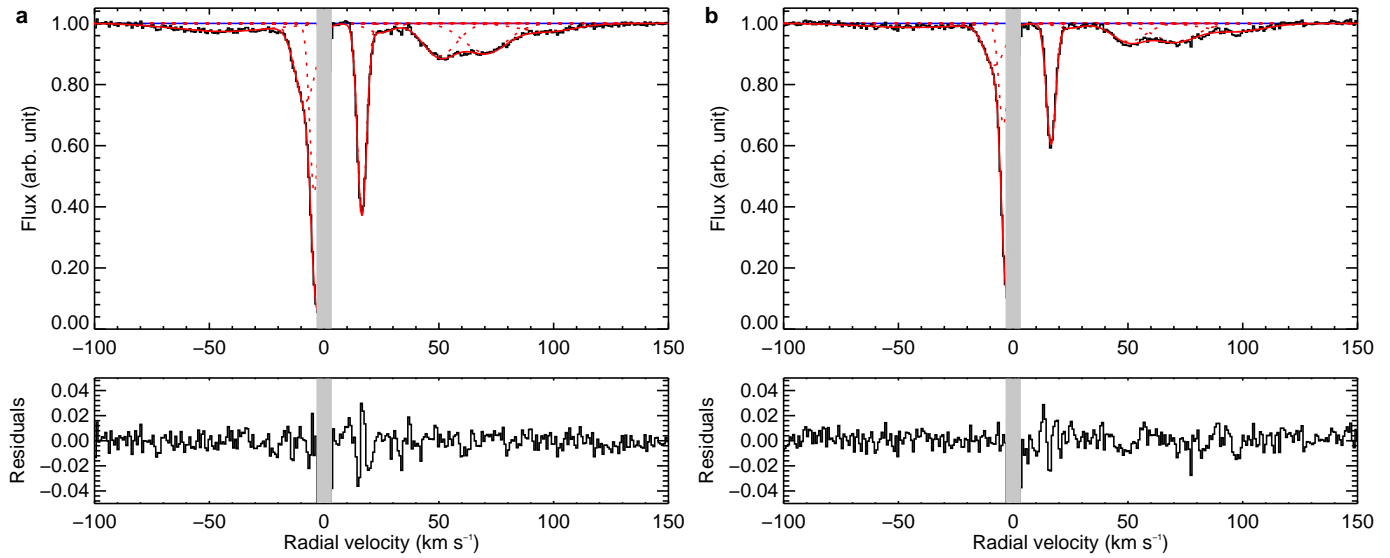
the circumstellar (CS) disk. The sharpness of the Na I D1 and D2 lines and the steadiness of this circumstellar feature in all spectra confirm the stability of HARPS on a timescale of years. The narrow absorption lines seen in most of the spectra and not in the calculated reference spectrum are due to atmospheric water.



**Extended Data Figure 2 | The Ca II reference spectrum of  $\beta$ Pictoris.** **a**, Ca II K-line spectrum (3,933.66 Å). **b**, Ca II H-line spectrum (3,968.47 Å). It shows the superposition of all the Ca II spectra of  $\beta$  Pic (black dots) compared with the stellar reference spectrum (red line). The stable circumstellar (CS) line is

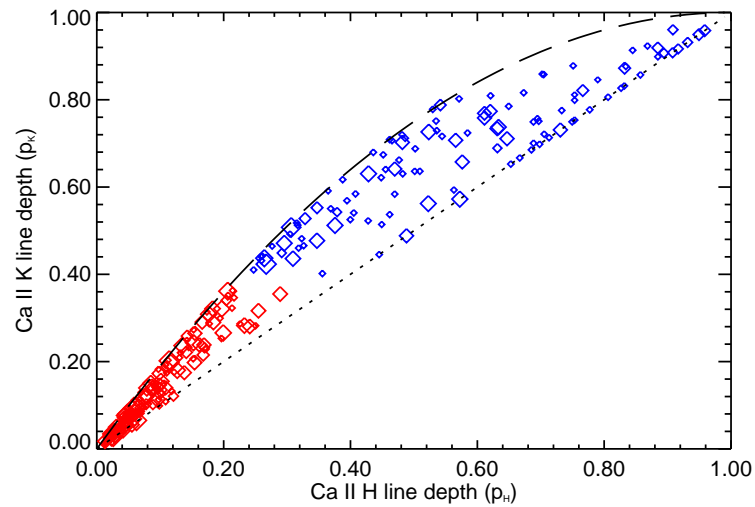
centred at the star's radial velocity. Variable absorption features are revealed by their diffuse shapes with respect to the dark upper envelop of the cloud of points. The predominance of redshifted absorption features is clearly visible. A small interstellar (IS) line is noticeable on the left of the circumstellar line<sup>24</sup>.





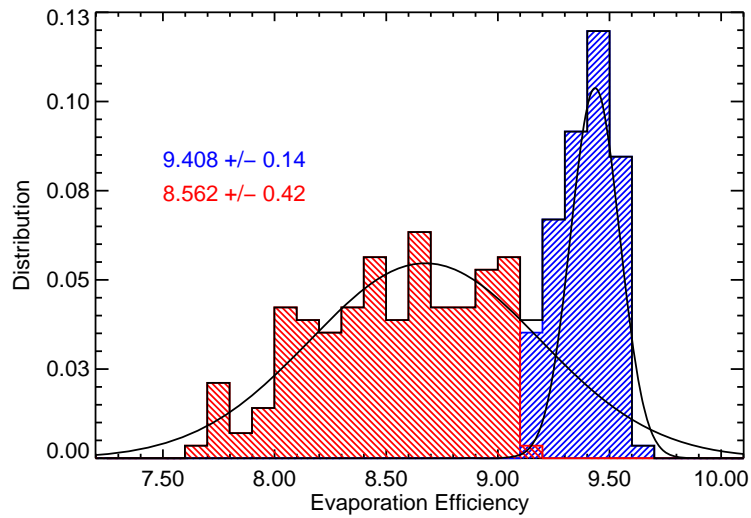
**Extended Data Figure 3 | A typical fitted Ca II normalized spectrum.** **a**, Ca II K line normalized spectrum. **b**, Ca II H line normalized spectrum. The Ca II normalized spectrum (black line) is obtained through the division of the corresponding regular spectrum collected on the 27 October 2009 (Fig. 1) by the reference spectrum plotted in Extended Data Fig. 2. Radial velocities are given with respect to the stellar rest frame. The fit of each feature detected is detailed with red dashed lines, and their superposition with a solid red line. The

bottom panels show the residuals of the fit. The grey shading indicates the  $\pm 3 \text{ km s}^{-1}$  excluded CS region, where variable absorption features caused by exocomets are not resolved from the circumstellar line. This spectrum presents all types of variable absorption features: a broad and shallow absorption at large radial velocity ( $\pm 50 \text{ km s}^{-1}$ ) and a sharp and deep absorption at small radial velocity ( $\sim 20 \text{ km s}^{-1}$ ).



**Extended Data Figure 4 | Diagram of the Ca II line depths.** Plot of the Ca II K line depth,  $p_K$ , as a function of the Ca II H line depth,  $p_H$ , for the 252 independent absorption features with  $\alpha < 1$  caused by individual transiting comets observed between 2003 and 2011. Using  $k$ -mean cluster analysis of these line depth measurements, two populations of exocomets show up: the 147 exocomets of population S generates the shallow absorption lines with

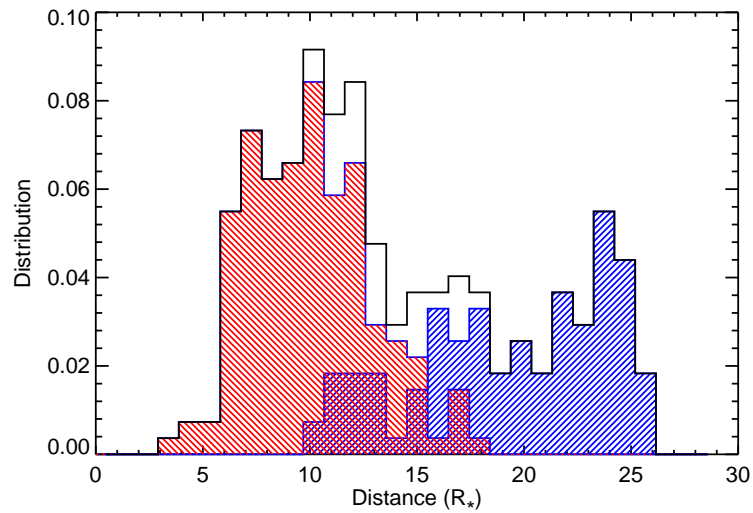
$p_K < 0.4$  (in red) and the 105 exocomets of population D generates the deep absorption lines with  $p_K > 0.4$  (in blue). The dotted line represents the full saturation limit  $p_K = p_H$  and the dashed line represents the  $\alpha = 1$  limit, corresponding to cometary cloud with a projected area greater than the stellar disk area.



**Extended Data Figure 5 | Histogram of the evaporation efficiency of transiting exocomets.** The histogram of  $\eta$ , the evaporation efficiencies (in black), shows a clear bimodal distribution: population S (in red) is centred on  $\eta_S = 8.6 \pm 0.4$ , while population D (in blue) is centred on  $\eta_D = 9.4 \pm 0.1$ .

The solid black histogram represents the distribution of evaporation efficiency for the 252 observed exocomets with  $\alpha < 1$ . The two Gaussian curves are obtained by fitting this histogram with the sum of two Gaussians.





**Extended Data Figure 6 | Histogram of the distances between ̢ Pic and the exocomets at the time of transit.** The 105 comets of population D (in blue) are located further away from the star than the 147 comets of

population S (in red), with  $d_D \approx 19 \pm 4 R_*$  and  $d_S \approx 10 \pm 3 R_*$ . Distances are expressed in units of stellar radius ( $R_*$ ). The solid black line represents the distribution of distances for the whole sample of observed exocomets.

# Characterizing and predicting the magnetic environment leading to solar eruptions

Tahar Amari<sup>1</sup>, Aurélien Canou<sup>1</sup> & Jean-Jacques Aly<sup>2</sup>

**The physical mechanism responsible for coronal mass ejections has been uncertain for many years, in large part because of the difficulty of knowing the three-dimensional magnetic field in the low corona<sup>1</sup>. Two possible models have emerged. In the first, a twisted flux rope moves out of equilibrium or becomes unstable, and the subsequent reconnection then powers the ejection<sup>2–5</sup>. In the second, a new flux rope forms as a result of the reconnection of the magnetic lines of an arcade (a group of arches of field lines) during the eruption itself<sup>6</sup>. Observational support for both mechanisms has been claimed<sup>7–9</sup>. Here we report modelling which demonstrates that twisted flux ropes lead to the ejection, in support of the first model. After seeing a coronal mass ejection, we use the observed photospheric magnetic field in that region from four days earlier as a boundary condition to determine the magnetic field configuration. The field evolves slowly before the eruption, such that it can be treated effectively as a static solution. We find that on the fourth day a flux rope forms and grows (increasing its free energy). This solution then becomes the initial condition as we let the model evolve dynamically under conditions driven by photospheric changes (such as flux cancellation). When the magnetic energy stored in the configuration is too high, no equilibrium is possible and the flux rope is ‘squeezed’ upwards. The subsequent reconnection drives a mass ejection.**

Coronal mass ejections (CMEs) are large-scale eruptive events in the solar atmosphere that could have impact<sup>10</sup> on satellites and ground-based power generation. Theoretical models<sup>11,12</sup> of their origins use a specific coronal configuration whose evolution is computed from a given initial state. A global disruption leading to the ejection of a twisted flux rope identified with a CME—through an overlying arcade—is then found to occur. The rope may be present in the initial state as a stable equilibrium structure; as an unstable or nearly unstable equilibrium structure that evolves freely at later times<sup>13,14</sup>; or as a subphotospheric structure that is forced by buoyancy to emerge through the solar surface into the corona, where it expands<sup>15,16</sup>. Alternatively, the initial state may simply be one or more arcades evolving through photospheric shearing motions, the rope being produced during the disruption itself<sup>6</sup>.

The following method can be used to determine which, if either, of these situations describes an actual eruptive event. In a first step, consider the pre-eruptive phase during which the coronal magnetic field can be assumed to evolve slowly through a sequence of equilibrium force-free configurations. Under this quite reasonable assumption, the field evolution can be computed from successive measurements made at the photospheric level, and the presence or absence of a twisted flux rope can be assessed. The problem is, however, very difficult, and safe conclusions can be reached only using a very accurate numerical code suited for force-free magnetic fields as well as high-resolution, low-noise measurements. In a second step, the configuration obtained slightly before the eruption is used as the initial state in a dynamical magnetohydrodynamics (MHD) code along with boundary conditions mimicking the physical photospheric processes that makes the coronal field evolve. It is then possible to look for a disruption of the configuration having properties comparable to the observed ones. It is worth noticing that this

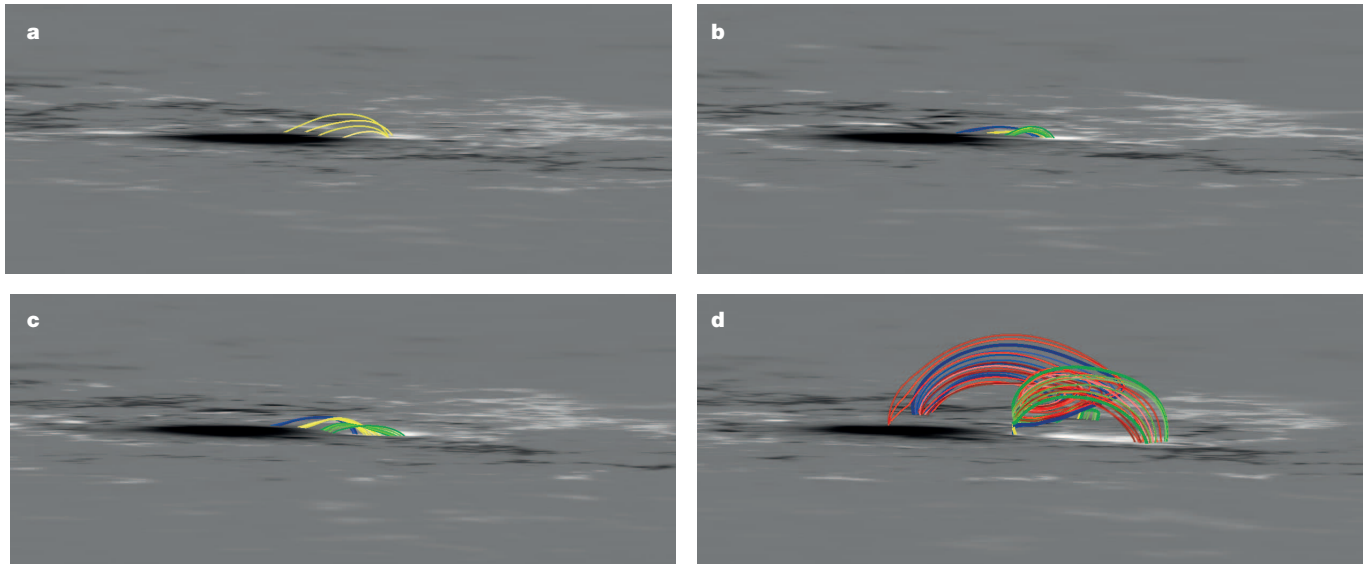
method, once successfully tested against well-documented past events, could be applied to forecast the eruption of an active region (the aim of studying space weather), the eruptive power at any given time being predicted by reconstructing the field at that time and using it in a dynamical code as indicated above.

Here we apply this method to US National Oceanic and Atmospheric Administration (NOAA) active region 10930 (AR10930), which crossed the solar disk during the first half of December 2006. This choice is motivated by two factors. First, the region clearly exhibits most of the features usually associated with eruptive behaviour. Second, the spectropolarimeter of the Solar Optical Telescope (SOT) on board the satellite Hinode<sup>17</sup> has provided a series of high-precision measurements of the photospheric field around the time of the eruption. This prompted many groups to study this region, but hitherto none of them has provided the complete, fully data-driven picture (including both the pre-eruptive and the eruptive phases) that we are seeking here. Most of them<sup>18–21</sup> (see Methods for supplementary references) have concentrated on the reconstruction problem, without showing unambiguously the presence of the pre-eruptive rope. As for the dynamical evolution of the region, an MHD computation leading to an eruption has been proposed<sup>22</sup>, but it starts from an initial state in which a twisted flux tube with adequate properties has been forced to emerge into the region rather than being obtained from a reconstruction based on the only observational magnetic data (a similar calculation, but with the tube introduced by the insertion method, was recently done for another region<sup>14</sup>). And the mechanism of the eruption has been discussed by comparing some observational features with the results of a simple theoretical model in which a twisted rope is forced to erupt into a highly sheared arcade<sup>23</sup>.

To set the stage, we note that the evolution of NOAA AR10930 was associated with the emergence of a positive-polarity sunspot<sup>24</sup>. This complex process involved translational and apparent rotational motions of the sunspot of the southern region, and led to the elongation of the spots, which is the signature of an emerging twisted rope<sup>25</sup> (Supplementary Video 1). A series of flares were observed during the transit of AR10930. An X3.4 two-ribbon flare occurred on December 13 from 2:14 UT to 2:57 UT, with its peak in soft X-rays at 2:40 UT. It had brightenings that started around the magnetic polarity inversion line between the two main spots, which then moved apart. It was followed by a halo-like CME with the potential to create a geomagnetic storm, which was observed using the C2 coronagraph on the Large Angle and Spectrometric Coronagraph Experiment on board NASA’s Solar and Heliospheric Observatory (SOHO) at 2:54 UT. The total energy (kinetic plus gravitational) released during this major event was estimated<sup>26</sup> to be  $(1.4\text{--}4.5) \times 10^{32}$  erg.

We compute the magnetic environment above AR10930 from December 9 to December 12 at 20:30 UT by using our extrapolation code XTRAPOL<sup>27</sup>, with the required boundary conditions being provided by vector magnetograms taken from the scans of the Hinode/SOT spectropolarimeter<sup>17</sup>. Several important features appear in the reconstructed states. As the evolution proceeds from day D–4 (four days before the eruption) to day D (the date of the eruption) as a consequence of flux emergence, the field suffers a change from an arcade-like topology

<sup>1</sup>Centre de Physique Théorique, Ecole Polytechnique, CNRS, F-91128 Palaiseau Cedex, France. <sup>2</sup>AIM, Unité Mixte de Recherche CEA, CNRS, Université Paris VII, UMR 7158, Centre d’Etudes de Saclay, F-91191 Gif sur Yvette Cedex, France.



**Figure 1 | Magnetic field evolution during the days before major eruption day D.** Selected field lines of some magnetic configurations obtained by reconstructing the coronal magnetic field around the eruption site using vector magnetograms from Hinode/SOT and a static MHD model. Energy builds up slowly, with field lines shearing several days before the eruption, during the emergence of the positive-polarity sunspot (white) into the background of an

older, negative-polarity sunspot (black). **a**, Four days before the eruption (D-4); **b**, D-3; **c**, D-2; **d**, D-1. The magnetic field becomes progressively sheared in between the two spots, until a large twisted flux rope has formed on D-1. Colour code: yellow, sheared arcades; blue and green, J-shaped loops; red, largely twisted field lines.

to a rope-like one (Fig. 1 and Supplementary Video 2). On D-2, the model produces two J-shaped structures surrounding a highly sheared inner arcade with almost no twist (Fig. 1c), similar to those produced in idealized models driven by flux cancellation<sup>2-4</sup>. On D-1, we also use our new global reconstruction code MESHMHD<sup>28</sup> along with composite photospheric data from three instruments, which allows us to resolve both small- and large-scale features (Extended Data Fig. 1). The active region from which the eruption originates is shown to be well isolated and not interacting with neighbouring active regions. A flux rope with a twist greater than  $2\pi$  has formed. There is evidence that the tube has a hyperbolic structure, which is known from previous models<sup>29</sup> to be the signature of a slow flux and magnetic helicity transfer by magnetic reconnection from the photosphere to the twisted rope (Extended Data Fig. 2). Some magnetic helicity appears to have been redistributed into the rope, resulting in an increase of its twist during the period from D-2 to D-1. The structure of the electric currents flowing along it on D-1 is shown in Extended Data Fig. 3.

The accuracy of our computed configurations is attested to by the very good agreement existing between some of their characteristic features and coronal observations. In particular, the magnetic lines of the twisted rope (which coincide with the lines of the current density in our force-free reconstructions) exhibit dips at the precise location of the filament of cool material observed in H $\alpha$  emission using the spectroheliograph at the Paris-Meudon Observatory (Fig. 2c, d), and they are well aligned with both the X-ray data from the Hinode X-ray Telescope (XRT) (Fig. 2b) and the extreme-ultraviolet data from the SOHO Extreme Ultraviolet Imaging Telescope (EIT) (Extended Data Fig. 4). The latter also show a good alignment with the magnetic lines of some outer parts of the region. Furthermore, the series of reconstructions made from D-4 to D-1 show that the rope is forming by emergence from below the photosphere as the southern spot of the observed active region is emerging, with a very good correlation with the observed photospheric tongues (Fig. 2a and Supplementary Video 2).

During the four modelled days, the magnetic energy increases considerably. Up to four days before the eruption, the amount of accumulated free energy—that is, the part of the energy above that of the current-free magnetic field—is too low to power it. At D-2, the free energy enters the ‘possible’ zone of eruption but is still close to the lower-bound estimate.

On D-1, the free magnetic energy becomes large enough to power the observed major eruption (Fig. 3a). A few hours before the eruption, the energy is close to that of a particular partially open field,  $B_{so}$ , whose importance has been stressed in previous idealized simulations<sup>4</sup> (Methods). Magnetostatic equilibrium is no longer possible when the energy of  $B_{so}$  is exceeded, a phenomenon known as catastrophic loss of equilibrium.

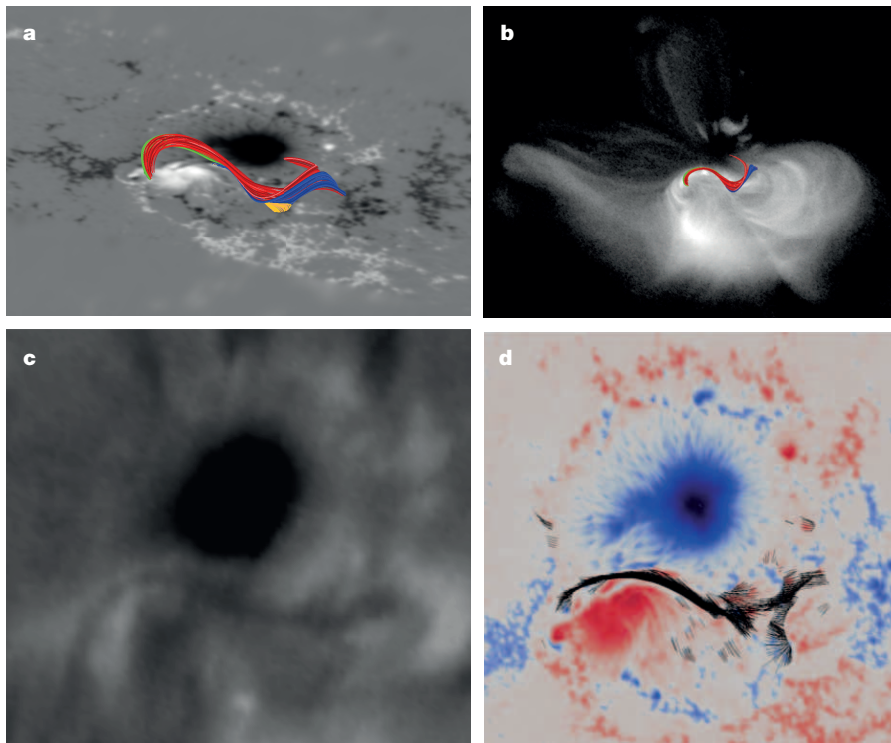
We take the low-corona configurations computed on D-2 and D-1, respectively, as initial conditions of our numerical dynamical MHD code METEOSOL<sup>2-4,30</sup>. The static solution evolves under the effects of a flux decrease<sup>2</sup> (associated with flux cancellation), gas motion characteristic of a sunspot ‘moat flow’<sup>4</sup>, or photospheric turbulent diffusion<sup>3,5</sup>.

We find that the magnetic configuration corresponding to the December 11, 10:00 UT (D-2) data does not lead to a major disruption, but rather to relaxation to a new equilibrium. Pushing the evolution further, however, results in the formation of a small twisted rope that eventually suffers a small, confined eruption. We do not get a major disruption because not enough flux has been transferred to the rope at D-2, with the available free energy staying well below the loss of equilibrium threshold. This shows that even at an early stage during the build-up of the active region the configuration has already acquired some eruptive potential (even if small).

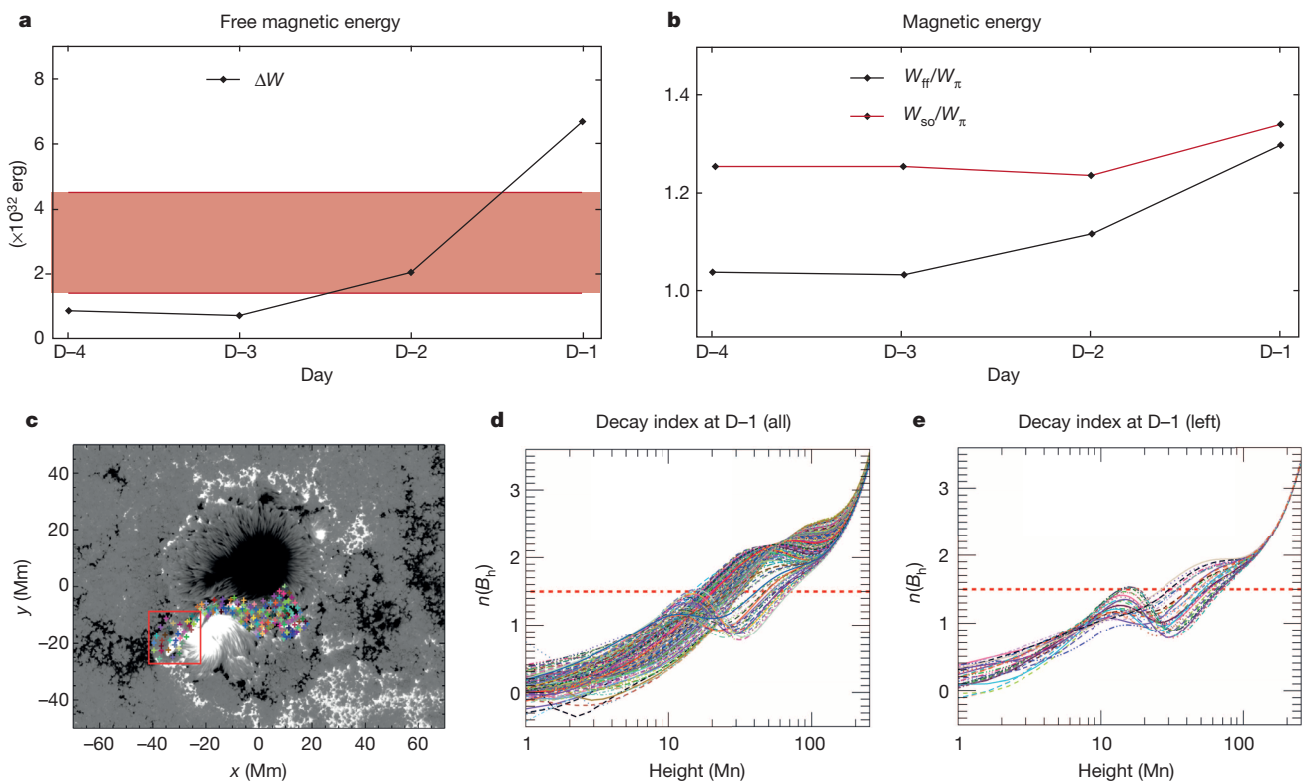
The situation is different when the evolution is computed with the data of D-1 at 20:30 UT. The twisted rope has already formed and contains a large amount of flux and electric current, and it is found to erupt, with the magnetic lines exhibiting a shape very similar to that generally reported when an eruption is observed on the solar limb (Fig. 4c, d). The overall configuration suffers a major disruption (Fig. 4a, b), which we argue explains the occurrence of the eruption of December 13, 2:40 UT. At the time of disruption, the magnetic energy of the configuration becomes of the same order as the critical energy (that of  $B_{so}$ ). We interpret this as meaning that the field has suffered a loss of equilibrium, where the magnetic tension force no longer balances the magnetic pressure force.

The onset of an eruption has been proposed<sup>5,13,14</sup> to occur when a rope reaches an altitude at which an index  $n$  characterizing the vertical decay of the horizontal component of the current-free field exceeds a critical value  $n_c \approx 3/2$  (the ‘torus instability criterion’). We have computed  $n$  for our configuration of day D-1 and found that  $n$  is close to  $n_c$  (Fig. 3c-e).



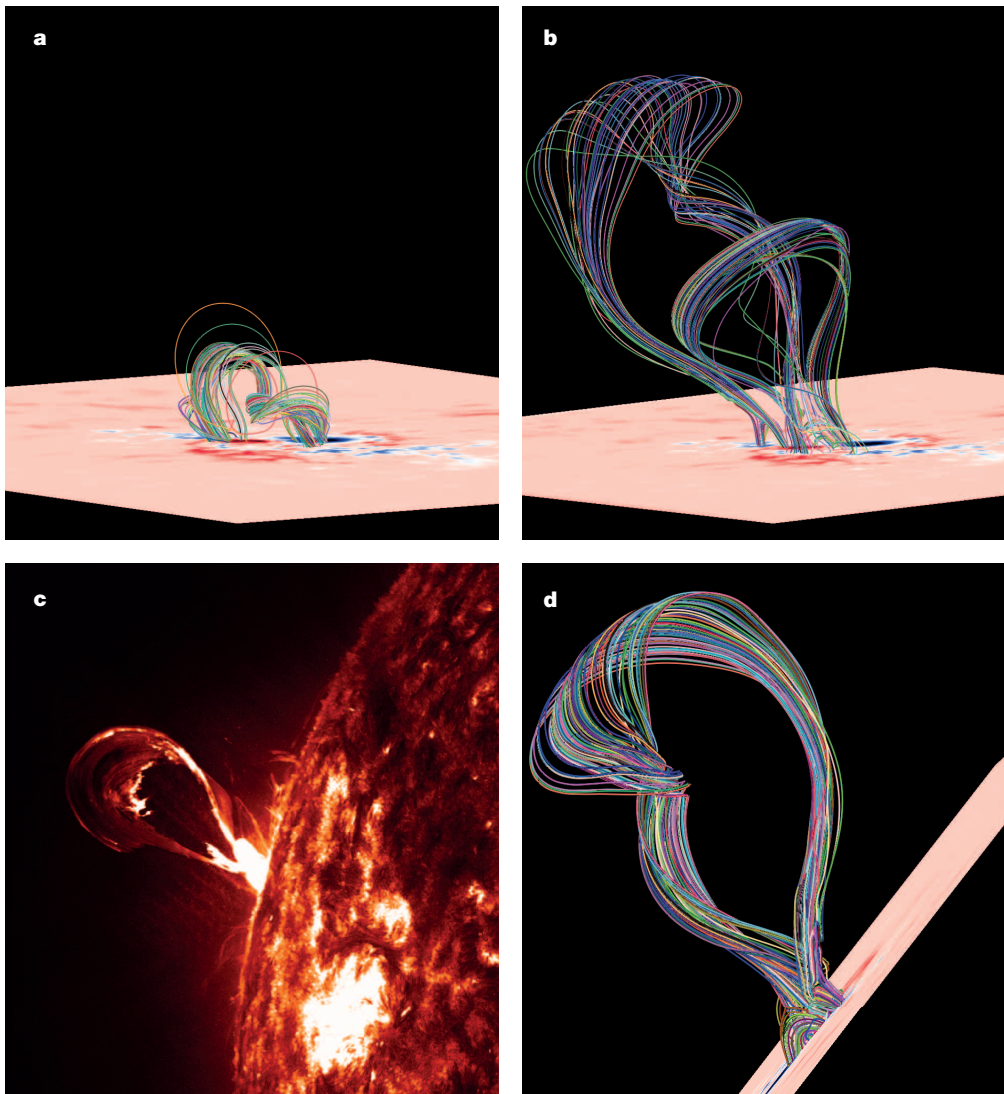


**Figure 2 | Twisted flux rope before the major eruption.** Selected field lines of the reconstructed magnetic configuration of December 12, 20:30 UT (D-1), with the same colour code as in Fig. 1. **a**, A large rope consisting of several components sits between the two spots and is seen to have accumulated a large amount of twist (about  $2.25\pi$ ). The hyperbolic nature of the rope (field lines bifurcating with an X-type topology) is detailed in Extended Data Fig. 2. **b**, Good agreement of the shape of some computed field lines with X-ray data from Hinode/XRT. **c**, H $\alpha$  data from the spectroheliograph at the Paris-Meudon Observatory reveals that a filament (darker) extends in the atmosphere between the two spots. **d**, The filament shown in **c** coincides with the locations of the dips in the computed magnetic field (shown as black segments and seen from the same vantage point as in **c**) where cool material can sit and be supported against gravity by the magnetic force.



**Figure 3 | Accumulation of magnetic energy.** **a**, Evolution of the free magnetic energy (black curve) during the four days before the major eruption. The actual energy of the major eruption lies in the red zone, defined by lower and upper limits estimated from the observations. **b**, Magnetic energy  $W_{\text{ff}}$  of the configuration (black curve) and theoretical magnetic energy upper bound  $W_{\text{so}}$ , beyond which equilibrium is no longer possible (red curve), in units of the magnetic energy  $W_{\pi}$  of the (minimum-energy) current-free magnetic configuration having the same  $B_z$  distribution on the photosphere. During the last day (D-1), the magnetic energy of the configuration comes closer to this

bound. **c**, Normal component of the magnetic field at D-1. The coloured crosses indicate a set of points selected along the photospheric projection of the twisted rope. **d**, Variations with altitude of the index  $n$  characterizing the decay of the horizontal component of the current-free magnetic field  $B_h$  above each of the coloured points in **c**. The red dashed line corresponds to the critical value  $n_c \approx 3/2$ , beyond which the rope is ejected according to the torus instability criterion. **e**, Variations in  $n$  for a subset of points lying inside the red rectangle in **c**, which is located on the left side of the southern spot.



**Figure 4 | Evolution and eruption of the twisted flux rope.** Selected field lines of some configurations obtained after an evolution driven by photospheric changes, with the configuration of December 12, 20:30 UT being taken as the initial condition. **a**, In accordance with Fig. 3b, at around 20:30 UT the configuration is almost out of equilibrium and the rope rapidly rises. **b**, Sometime later, it erupts through the arcs that were able to confine it several hours before. **c**, Observed typical shape exhibited by eruptions on the solar limb (image taken at a wavelength of 304 Å by SDO/AIA on 2012 October 14, at 2:32 UT; courtesy of NASA/SDO and the AIA, EVE and HMI science teams). **d**, Modelled appearance of the major eruption of December 13, projected onto the solar limb, showing agreement with the typical shape shown in **c** (for convenience, the arcs have been removed).

There is then an agreement between our energetic onset criterion and the torus instability one. Moreover,  $n$  is closer to  $n_c$  in the region near the left side of the southern spot, which might explain the asymmetric ejection of the rope whose eruption is stronger above that region (Fig. 3e). Finally, during the MHD phase of the rising of the rope, most of it eventually reaches a height at which  $n = n_c$ . It has been suggested<sup>14</sup> that the onset of an eruption may be also characterized in terms of a flux criterion, with the rope being ejected once the ratio of its axial flux to the flux through the main part of the active region becomes too large. We find that this parameter is equal to 15.3% at D = 1.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 22 July; accepted 29 August 2014.**

- Cargill, P. Coronal magnetism: difficulties and prospects. *Space Sci. Rev.* **144**, 413–421 (2009).
- Amari, T., Luciani, J.-F., Mikic, Z. & Linker, J. A twisted flux rope model for coronal mass ejection and two-ribbon flares. *Astrophys. J.* **529**, L49–L52 (2000).
- Amari, T., Luciani, J.-F., Aly, J.-J., Mikic, Z. & Linker, J. Coronal mass ejection: initiation, helicity and flux ropes. II. Turbulent diffusion driven evolution. *Astrophys. J.* **595**, 1231–1250 (2003).
- Amari, T., Aly, J.-J., Luciani, J.-F., Mikic, Z. & Linker, J. Coronal mass ejection initiation by converging photospheric flows: toward a realistic model. *Astrophys. J.* **742**, L27 (2011).
- Aulanier, G., Török, T., Démoulin, P. & DeLuca, E. E. Formation of torus-unstable flux ropes and electric currents in erupting sigmoids. *Astrophys. J.* **708**, 314–333 (2010).
- Antiochos, S. K., DeVore, C. R. & Klimchuk, J. A. A model for coronal mass ejection. *Astrophys. J.* **510**, 485–493 (1999).
- Howard, T. A. & DeForest, C. E. Inner heliospheric flux rope evolution via imaging of coronal mass ejections. *Astrophys. J.* **746**, 64–76 (2012).
- Cheng, X. *et al.* Tracking the evolution of a coherent magnetic flux rope continuously from the inner to the outer corona. *Astrophys. J.* **780**, 28 (2014).
- Kumar, P. & Innes, D. E. Multiwavelength observations of an eruptive flare: evidence for blast wave and break out. *Sol. Phys.* **288**, 255–268 (2013).
- Hapgood, M. Prepare for the coming space weather storm. *Nature* **484**, 311 (2012).
- Forbes, T. G. *et al.* CME theory and model. *Space Sci. Rev.* **123**, 251–302 (2006).
- Shibata, K. & Magara, T. Solar flares: magnetohydrodynamic processes. *Living Rev. Sol. Phys.* **8**, 6 (2011).
- Kliem, B. & Török, T. Torus instability. *Phys. Rev. Lett.* **96**, 255002 (2006).
- Kliem, B. *et al.* Magnetohydrodynamic modeling of the solar eruption on 2010 April 8. *Astrophys. J.* **779**, 129 (2013).
- Fan, Y. The emergence of a twisted flux tube into the solar atmosphere: sunspot rotations and the formation of a coronal flux rope. *Astrophys. J.* **697**, 1529–1542 (2009).
- Roussev, I. I. *et al.* Explaining fast ejection of plasma and exotic X-ray emission from the solar corona. *Nature Phys.* **8**, 845–849 (2012).
- Kosugi, T. *et al.* The Hinode (Solar-B) mission: an overview. *Sol. Phys.* **243**, 3–17 (2007).
- Schrijver, C. J. *et al.* Nonlinear force-free field modeling of a solar active region around the time of a major flare and coronal mass ejection. *Astrophys. J.* **675**, 1637–1644 (2008).
- Guo, Y., Ding, M. D., Wiegmann, T. & Li, H. 3D magnetic field configuration of the 2006 December 13 flare extrapolated with the optimization method. *Astrophys. J.* **679**, 1629–1635 (2008).
- Jing, J., Wiegmann, T., Suematsu, Y., Kubo, M. & Wang, H. Changes of magnetic structure in three dimensions associated with the X3.4 flare of 2006 December 13. *Astrophys. J.* **676**, L81–L84 (2008).

21. Inoue, S., Kusano, K., Magara, T., Shiota, D. & Yamamoto, T. T. Twist and connectivity of magnetic field lines in the solar active region NOAA 10930. *Astrophys. J.* **738**, 161 (2011).
22. Fan, Y. A magnetohydrodynamic model of the 2006 December 13 eruptive flare. *Astrophys. J.* **740**, 68 (2011).
23. Kusano, K. *et al.* Magnetic structures triggering solar flares and coronal mass ejections. *Astrophys. J.* **760**, 31 (2012).
24. Min, S. & Chae, J. The rotating sunspot in AR 10930. *Sol. Phys.* **258**, 203–217 (2009).
25. Luoni, M. L. *et al.* Twisted flux tube emergence evidenced in longitudinal magnetograms: magnetic tongues. *Sol. Phys.* **270**, 45–74 (2011).
26. Ravindra, B. & Howard, T. A. Comparison of energies between eruptive phenomena and magnetic field in AR 10930. *Bull. Astron. Soc. India* **38**, 147–163 (2010).
27. Amari, T., Boulmezaoud, T. Z. & Aly, J.-J. Well posed reconstruction of the solar coronal magnetic field. *Astron. Astrophys.* **446**, 691–705 (2006).
28. Amari, T. *et al.* in *Proc. ASP Conf. Ser. 459* (eds Pogorelov, N. V., Font, J. A., Audit, E. & Zank, G. P.) 189 (Astronomical Society of the Pacific, 2012).
29. Titov, V. S. & Hornig, G. Magnetic connectivity of coronal fields: geometrical versus topological description. *Adv. Space Res.* **29**, 1087–1092 (2002).
30. Amari, T., Luciani, J. F. & Joly, P. A preconditioned semi implicit scheme for magnetohydrodynamics equations. *SIAM J. Sci. Comput.* **21**, 970–986 (1999).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** The numerical simulations described in this paper were performed on the IBM x3750 of the IDRIS institute of the Centre National de la Recherche Scientifique (CNRS). We thank the Centre National d'Etudes Spatiales (CNES) for its financial support. Hinode is a Japanese mission developed and launched by ISAS/JAXA, with NAOJ as domestic partner and NASA and STFC (UK) as international partners. It is operated by these agencies in co-operation with ESA and NSC (Norway). H $\alpha$  data used in this study were provided by Paris-Meudon Observatory, X-ray data came from Hinode/XRT and extreme-ultraviolet data came from SOHO/EIT.

**Author Contributions** T.A. and A.C. planned and performed the various calculations, and discussed the analysis with J.-J.A. The manuscript was written by T.A. and J.-J.A. with feedback from A.C.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.A. ([amari@cph.tpolytechnique.fr](mailto:amari@cph.tpolytechnique.fr)).



## METHODS

The basic strategy that we have adopted is the following: first, we model the pre-eruptive phase through a series of reconstructions of the field (assumed to be in force-free equilibrium), with the needed boundary conditions being computed from the observational data. Second, we predict the evolution of the field during the eruptive phase by using the last obtained equilibrium as the initial condition of a simulation done with a full MHD code. The physical results we get depend on both the quality of the data used and the mathematical efficiency of the method.

**Magnetic environment model at equilibrium.** The magnetic field  $\mathbf{B}$  can be measured only in the photosphere, but we need a coronal value. During the occurrence of an eruptive event,  $\mathbf{B}$  and the highly conducting, low- $\beta$  plasma in which it is embedded evolve very slowly.  $\mathbf{B}$  can thus be considered as being at each time  $t$  in an equilibrium force-free state in which it obeys the set of equations<sup>31</sup>

$$\nabla \times \mathbf{B} = \alpha(\mathbf{r})\mathbf{B} \quad (1)$$

$$\nabla \cdot \mathbf{B} = 0 \quad (2)$$

$$\mathbf{B} \cdot \nabla \alpha(\mathbf{r}) = 0 \quad (3)$$

where  $\alpha(\mathbf{r})$  is some scalar function of position  $\mathbf{r}$ . Equation (3) (which is a consequence of equations (1) and (2)) implies that  $\alpha$  keeps a constant value along any field line. One is thus led to set up the ‘reconstruction problem’: to determine in the part  $\Omega$  of the corona above an active region a magnetic field  $\mathbf{B}$  satisfying as closely as possible the following conditions: (1) it is a finite-energy, force-free magnetic field in  $\Omega$  and (2) it matches on the lower boundary  $S$  of  $\Omega$  the value  $\mathbf{B}^{\text{phot}}$  provided by the measurements done at the photospheric level<sup>31</sup>. The reason why we added the restrictive qualification ‘as closely as possible’ is that there is generally no field  $\mathbf{B}$  satisfying conditions (1) and (2) simultaneously. The raw data can be ‘preprocessed’<sup>32</sup> to try to diminish the incompatibility between both requirements, but it cannot be totally suppressed. The best one can do is to set up a resolution scheme which is well-posed, in the sense that it leads to a unique solution that is stable with respect to small changes in the data.

Several methods<sup>31</sup> have been developed to treat the reconstruction problem, each one using the data in a specific way and defining a particular environment model, and it is important when one studies a particular active region to compare the results furnished by various approaches<sup>18</sup>. One should also compare these results with characteristic observational features of the region (like the presence of a filament at some particular location).

To reconstruct the coronal field, we use a Grad–Rubin type method that solves a mixed hyperbolic–elliptic boundary-value problem for the force-free function  $\alpha$  and for the field  $\mathbf{B}$  (ref. 27). This method is based on rigorous mathematical grounds and has been proven to be well posed<sup>33</sup>. As for the boundary conditions, one has to fix the value  $g$  of the normal component of the field on the whole boundary  $S$  and the value  $\lambda$  of  $\alpha$  on either the positive-polarity part  $S^+$  of  $S$  (where  $B_z > 0$ ) or on the negative-polarity part  $S^-$  (where  $B_z < 0$ ). The XTRAPOL code<sup>27</sup> solves this problem by means of the iterative Grad–Rubin scheme:

$$\mathbf{B}^{(k)} \cdot \nabla \alpha^{(k)} = 0 \quad \text{in } \Omega \quad (4)$$

$$\alpha^{(k)} = \lambda \quad \text{on } S^+ \text{ or } S^- \quad (5)$$

$$\nabla \times \mathbf{B}^{(k+1)} = \alpha^{(k)} \mathbf{B}^{(k)} \quad \text{in } \Omega \quad (6)$$

$$\nabla \cdot \mathbf{B}^{(k+1)} = 0 \quad \text{in } \Omega \quad (7)$$

$$\mathbf{B}_n^{(k+1)} = g \quad \text{on } S \quad (8)$$

The initial field  $\mathbf{B}^{(0)}$  is chosen to be the potential magnetic field  $\mathbf{B}_p$  with a normal component equal to  $g$  on  $S$  (by definition  $\mathbf{B}_p$  is current free, that is,  $\nabla \times \mathbf{B}_p = 0$ ). XTRAPOL uses a finite-difference method with a representation of  $\mathbf{B}$  based on a vector potential  $\mathbf{A}$  (with a convenient choice of gauge), which ensures that  $\nabla \cdot \mathbf{B} = 0$  to the accuracy of round-off errors. The elliptic part (equations (6)–(8)) for  $\mathbf{B}^{(k+1)}$  is solved through a positive-definite linear system, and the hyperbolic part (equations (4)–(5)) for  $\alpha^{(k)}$  is solved by transporting the values  $\lambda$  imposed on either  $S^+$  or  $S^-$  along the magnetic field lines of  $\mathbf{B}^{(k)}$ . The code uses the message passive interface library. It provides a solution even if the photospheric flux is not balanced.

**Magnetic data and boundary conditions.** The boundary conditions needed by our method,  $g$  and  $\lambda$ , are extracted from the photospheric vector magnetograms obtained from the scans of the spectropolarimeter of the Solar Optical Telescope<sup>34</sup> on board the satellite Hinode<sup>17</sup>. More precisely, we used level-2 data (available online at <http://sot.lmsal.com/data/sot/level2d/>) obtained by applying the MERLIN inversion code<sup>35</sup> to the Stokes parameters  $I$ ,  $Q$ ,  $U$  and  $V$  measured by this instrument. The well-known 180° ambiguity suffered by the transverse component of the magnetic field is resolved by using the minimum-energy code (MEO) based on a recent re-implementation<sup>36,37</sup> of the original method<sup>38</sup>. By using standard transformation formulae<sup>39</sup>, we have also converted the scales and the magnetic components from the observer frame to the Cartesian frame tangent to the solar surface and centred on the middle point of the vector magnetograms. We have chosen a sequence of four vector magnetograms covering the period from December 9, 10:00 UT, to December 12, 20:30 UT (the last available data before the eruption). Note that the raw data provided by the instrument are neither smoothed nor preprocessed.

The value of  $g$  is directly furnished by the data. That of  $\lambda$  is computed from the three components of the measured magnetic field  $\mathbf{B}^{\text{phot}}$  according to<sup>40</sup>

$$\lambda = \frac{4\pi}{c} \frac{j_z^{\text{phot}}}{B_z^{\text{phot}}} = \frac{(\nabla \times \mathbf{B}^{\text{phot}})_z}{B_z^{\text{phot}}}$$

where  $j_z^{\text{phot}}$  is the vertical component of the current density and  $c$  is the speed of light. To prevent unreliable values of  $\lambda$  near the polarity inversion line where  $B_z^{\text{phot}}$  is small,  $\lambda$  is set to zero if  $|B_z^{\text{phot}}|$  is below a particular value,  $B_z^{\text{cut}}$ . A similar cut-off is used for the intensity of the tangential component  $B_t^{\text{phot}}$  of  $\mathbf{B}$ . This avoids unreliable values of  $j_z^{\text{phot}}$  due to sudden variations of  $B_t^{\text{phot}}$  below the noise level. It should be noted that some improvements leading to non-zero values of  $\lambda$  closer to the polarity inversion line could be obtained by applying further smoothing and interpolation on the computational mesh.

**Magnetic environment properties.** Using these boundary conditions, we compute the solutions of our reconstruction problem with a numerical resolution of  $501 \times 331 \times 201$ . The quality of the results is evaluated by calculating the standard a-posteriori diagnostics, which are found to be much better than those characterizing previous reconstructions<sup>18–21,41,42</sup>. In particular, we obtain much smaller values for the angle between the magnetic field and the electric current density ( $3.73^\circ$  versus  $14.48^\circ$  (model<sup>18</sup>  $W_{pp}^+$ )), these values being computed from the actual diagnostic CWin, which deals with the sine of this angle) and for the functional  $L_f$  measuring the distance to a true equilibrium ( $0.07 \text{ G}^2 \text{ Mm}^{-2}$  versus  $2.27 \text{ G}^2 \text{ Mm}^{-2}$  in ref. 19). Owing to our specific discretization alluded to above, the difference is even larger for the functionals measuring the residual divergence of the computed magnetic field. We obtain  $L_d = 10^{-30} \text{ G}^2 \text{ Mm}^{-2}$  versus  $L_d = 1.15 \text{ G}^2 \text{ Mm}^{-2}$ , from ref. 19, where the reconstructions were made by the optimization method and the constraint  $\nabla \cdot \mathbf{B} = 0$  was not imposed a priori, the field being made only as ‘divergenceless’ as possible. Also, we obtain  $\langle |f_i| \rangle = 2.7 \times 10^{-17}$  versus  $3.6 \times 10^{-8}$  for the best model<sup>18</sup>  $W_{pp}^+$ , where a Grad–Rubin scheme is used, but with a discretization leading to a non-zero resolution dependent value of  $\nabla \cdot \mathbf{B}$  at each node. As another test, we have considered the difference between the measured values of  $\alpha$  at the two ends of any magnetic line, which in principle should vanish if the boundary data are strictly compatible with the existence of a force-free field in  $\Omega$ . We find that this quantity is relatively small for the field we compute despite the fact that our well-posed formulation uses only the values  $\lambda$  of  $\alpha$  in the positive-polarity region. It in fact takes relatively large values for the reconstructions based on the use of all three components of the magnetic field on the whole active region.

We also calculate the free magnetic energy  $\Delta W$  (that is, the difference between the magnetic energy of the solution and that of the associated current-free field  $\mathbf{B}_p$ ) stored in the equilibria. This is a very important quantity to be compared with the energy released during the following X-class eruption. For our last computed pre-flare configuration, we find  $\Delta W = 7 \times 10^{32} \text{ erg}$ , which is above the upper limit of the estimate  $((1.4\text{--}4.5) \times 10^{32} \text{ erg})$  of the energy released during the flare<sup>26</sup>. The good observational compatibility of this result can be taken as another indication of the quality of our reconstructions (in particular, we recall that having a low value of  $L_d$  is well known to be necessary for obtaining an accurate value of the energy<sup>43</sup>). Note that the energies previously found<sup>18,19</sup> ( $\Delta W = 5 \times 10^{32}$  and  $1.3 \times 10^{32} \text{ erg}$ ) are also compatible with the energy release estimates, with the second value<sup>19</sup> being close to the lower limit.

As noted above, it is important to reconstruct the coronal field by different methods and to check for the coherence of the results. We have also done reconstructions using a newly developed numerical code, MESHMHDP<sup>28</sup>, which solves the equilibrium equations (4)–(8) in the whole spherical corona. It is an adaptive unstructured (tetrahedral) mesh code, which includes in particular a different scheme for computing  $\alpha$ . This scheme solves a linear non-symmetric system using the GMRES algorithm<sup>44</sup>. The boundary conditions are provided by composite photospheric data from three instruments: a Hinode/SOT vector magnetogram at the active region scale, a SOHO/MDI full-disk longitudinal magnetogram and a SOLIS synoptic

map with a latitude–longitude resolution of  $1,368 \times 2,824$ . As it is based on an unstructured mesh, MESHMHD can introduce high resolution where it is needed in the region and around the flux rope, and lower resolution elsewhere. As a striking result, we found that the two codes detect the presence of a twisted flux rope and capture its bifurcation to different anchoring points on the east side (Extended Data Fig. 2).

**Prediction of the later evolution by using a dynamical MHD code.** During the eruptive phase, the coronal magnetic field evolves according to its own internal dynamics and can then no longer be determined from its photospheric values. But we can use our dynamical MHD code METEOSOL<sup>30</sup>, which solves the full system of MHD equations, to predict its evolution from the last computed pre-eruptive equilibrium state. We inject that state into the code as an initial condition and select boundary conditions able to describe the actual photospheric processes that force the field to evolve. In our previous theoretical studies<sup>3,4,45</sup>, we identified three such processes and showed that each of them can trigger a flare and a CME:

- (1) Partial cancellation near the polarity inversion line<sup>44</sup>.
- (2) Turbulent diffusion<sup>3,5,46</sup>, which is an important process permanently acting on the whole surface of the Sun, where it leads to the continuous dispersion of active regions. It can be characterized by a coefficient of turbulent diffusion  $\kappa_B = 10^{-3}$  or  $10^{-4}$ .
- (3) Plasma motions that diverge from the two spots of a bipolar region and converge towards the polarity inversion line in its vicinity, into which they transport a part of the magnetic flux<sup>4</sup>. These ‘moat flows’, which have been observed in an active region<sup>47,48</sup>, have their source just outside the penumbra, near the locus at which the Evershed flow disappears.

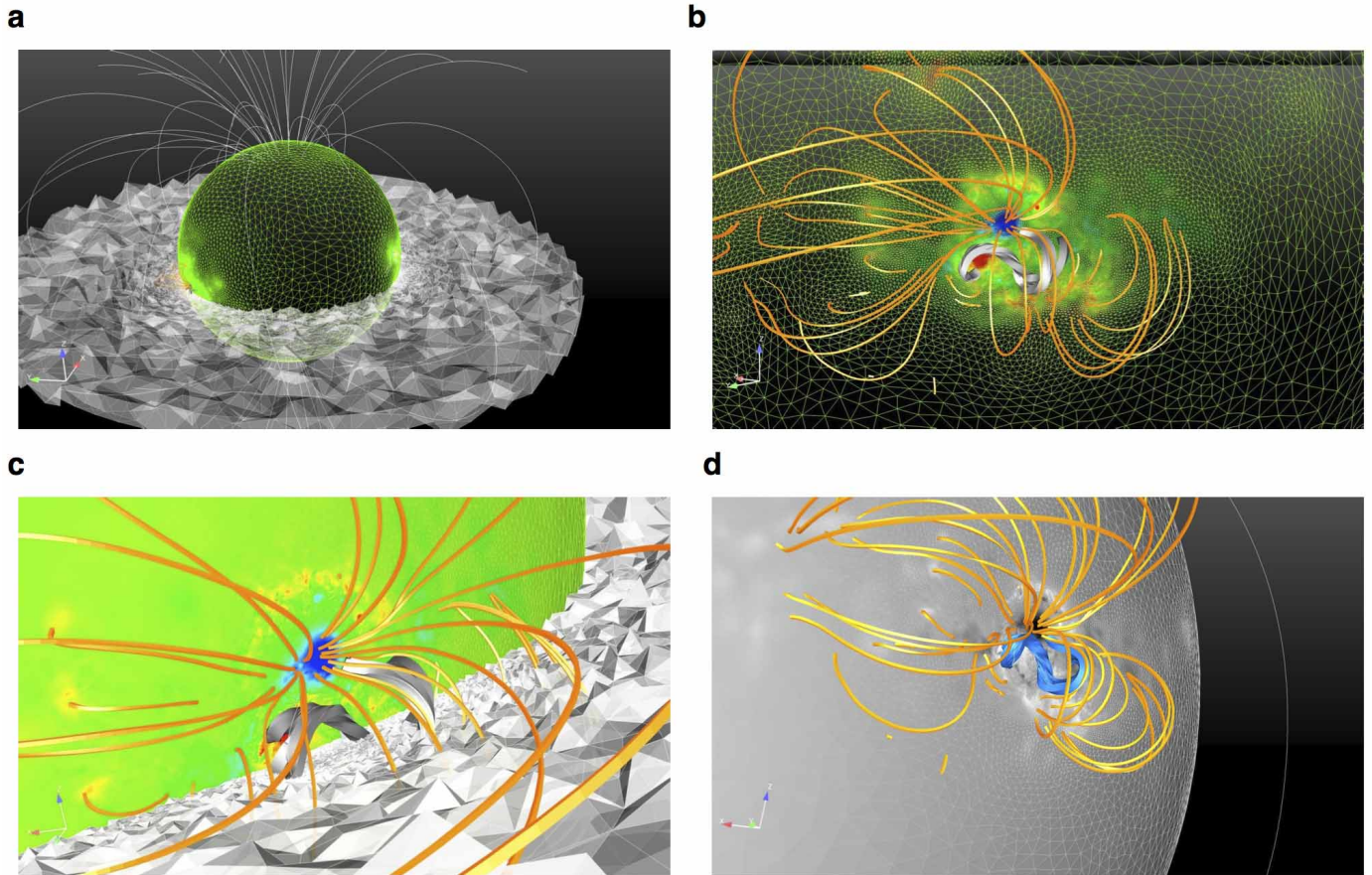
We successively compute three MHD evolutions, each starting from the same initial state and driven by one of the processes above. The processes are introduced into the code by requiring the tangential component of the electric field on S to assume a specific form.

The three types of boundary condition lead after a short interval to a major disruption. This is similar to what we obtained in the corresponding theoretical models previously developed. There is, however, an important difference: here the initial presence and properties of the twisted rope are no longer postulated, but are furnished by our environment model, which determines to a very good approximation the actual initial structure of the field above the active region. Reconstruction has shown the presence of a rope in an active region<sup>49</sup>, but this has been rare. A striking result of our modelling is that the eruption is due to a loss of equilibrium occurring when the energy of the configuration becomes of the order of the energy of the partially open field  $B_{so}$  defined below.

**The partially open field.** To define  $B_{so}$ , we first construct an additional magnetic field, the open field  $B_s$ , having the same normal component  $g$  on S as the evolving coronal field  $B$ . In contrast to the potential field  $B_p$ , whose most magnetic lines are closed (each one connects two points of S),  $B_s$  has all its field lines open (they connect S to the upper and lateral boundaries of the computational box). It is current free except on the surface separating outgoing lines from ingoing ones, across which it reverses. As with  $B_p$ ,  $B_s$  is in equilibrium in the sense that it does not exert any force on the plasma in which it is embedded. We can also construct equilibrium fields,  $B'_s$ , satisfying the same boundary condition and ‘interpolating’ between  $B_p$  and  $B_s$ . These fields are only partially open—they have both closed and open lines—and they are also current free except on a reversal surface. The field  $B_{so}$  is one of

these fields  $B'_s$ . It is selected by requiring that its open lines connect to S in the region where the electric currents are concentrated. This field was first introduced in an analytical theory<sup>31</sup> that describes the evolution of  $B$  as a sequence of equilibrium states, this quasi-static approximation being valid as long as the evolution is slow. It was argued on general grounds that  $W(B) \leq W(B_{so}) \leq W(B'_s)$  (with  $W(B)$  denoting the energy of  $B$ ), and that  $B$  starts experiencing a fast expansion leading to its partial opening when  $W(B)$  approaches  $W(B_{so})$ . Such fast evolution implies a breakdown of the quasi-static approximation and one needs to adopt the dynamical approach described in the previous item to get a proper description of it.

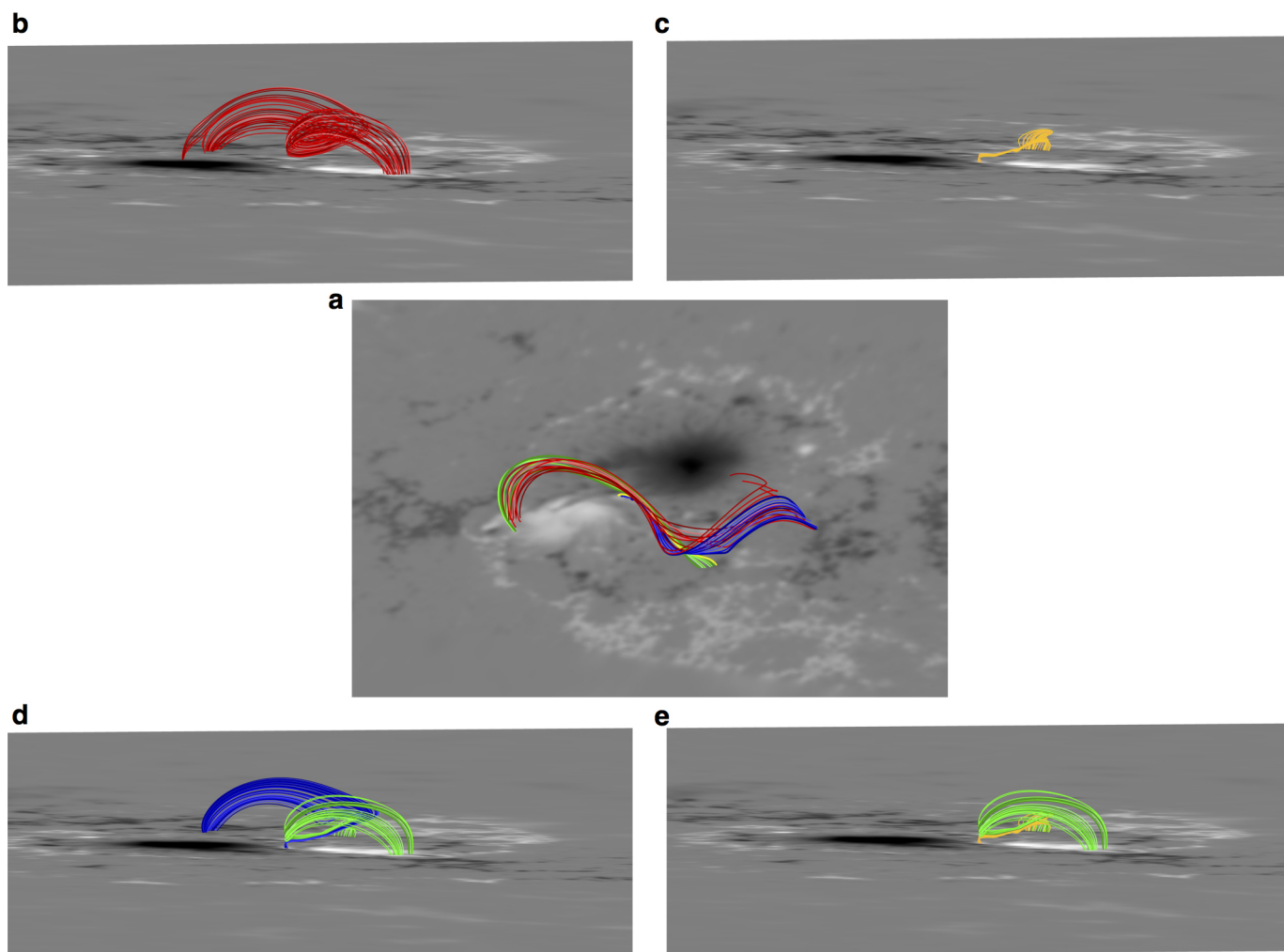
31. Aly, J.-J. & Amari, T. Structure and evolution of the solar coronal magnetic field. *Geophys. Astrophys. Fluid Dyn.* **101**, 249–287 (2007).
32. Wiegmann, T., Inhester, B. & Sakurai, T. Preprocessing of vector magnetograph data for a nonlinear force-free magnetic field reconstruction. *Sol. Phys.* **233**, 215–232 (2006).
33. Boulmezaoud, T. Z., & Amari, T. On the existence of non-linear force-free fields in three-dimensional domains. *Z. Angew. Math. Phys.* **51**, 942–967 (2000).
34. Tsuneta, S. *et al.* The Solar Optical Telescope for the Hinode mission: an overview. *Sol. Phys.* **249**, 167–196 (2008).
35. Skumanich, A. & Lites, B. W. Stokes profile analysis and vector magnetic fields. I - Inversion of photospheric lines. *Astrophys. J.* **322**, 473–482 (1987).
36. Leka, K. D. *et al.* Resolving the 180° ambiguity in solar vector magnetic field data: evaluating the effects of noise, spatial resolution, and method assumptions. *Sol. Phys.* **260**, 83–108 (2009).
37. Crouch, A. D., Barnes, G. & Leka, K. D. Resolving the azimuthal ambiguity in vector magnetogram data with the divergence-free condition: application to discrete data. *Sol. Phys.* **260**, 271–287 (2009).
38. Metcalf, T. R. Resolving the 180-degree ambiguity in vector magnetic field measurements: the ‘minimum’ energy solution. *Sol. Phys.* **155**, 235–242 (1994).
39. Allen Gary, G. & Hagyard, M. J. Transformation of vector magnetograms and the problems associated with the effects of perspective and the azimuthal ambiguity. *Sol. Phys.* **126**, 21–36 (1990).
40. Régnier, S., Amari, T. & Kersalé, E. 3D coronal magnetic field from vector magnetograms: non-constant-alpha force-free configuration of the active region NOAA 8151. *Astron. Astrophys.* **392**, 1119–1127 (2002).
41. Fan, Y.-L., Wang, H.-N., He, H. & Zhu, X.-S. Application of a data-driven simulation method to the reconstruction of the coronal magnetic field. *Res. Astron. Astrophys.* **12**, 563–572 (2012).
42. Inoue, S. *et al.* Buildup and release of magnetic twist during the X3.4 solar flare of 2006 December 13. *Astrophys. J.* **760**, 17 (2012).
43. Valori, G., Démoulin, P., Parlat, E. & Masson, S. Accuracy of magnetic energy computations. *Astron. Astrophys.* **553**, A38 (2013).
44. Saad, Y. & Schultz, M. H. GMRES: a generalized minimal residual algorithm for solving non symmetric linear systems. *SIAM J. Sci. Stat. Comput.* **7**, 856–869 (1986).
45. Amari, T., Aly, J.-J., Luciani, J.-F., Mikic, Z. & Linker, J. Coronal mass ejection initiation: on the nature of the flux cancellation model. *Astrophys. J.* **717**, L26 (2010).
46. Wang, Y. M., Sheeley, N. R. & Nash, A. G. A new solar cycle model including meridional circulation. *Astrophys. J.* **383**, 431–442 (1991).
47. Chae, J. *et al.* The formation of a prominence in active region NOAA 8668. I. SOHO/MDI observations of magnetic field evolution. *Astrophys. J.* **560**, 476–489 (2001).
48. Lites, B. W., Socas-Navarro, H., Skumanich, A. & Shimizu, T. Converging flows in the penumbra of a  $\delta$  sunspot. *Astrophys. J.* **575**, 1131–1143 (2002).
49. Yan, Y. *et al.* The magnetic rope structure and associated energetic processes in the 2000 July 14 solar flare. *Astrophys. J.* **551**, L115–L119 (2001).



**Extended Data Figure 1 | Another multiscale model.** Full-Sun magnetic configuration obtained using composite data set (Hinode/SOT and SOLIS synoptic map) and the state-of-the-art numerical code MESHMHD, which is a tetrahedral adaptive-mesh equilibrium code. Local and global scales are both accessible using very high-resolution data around the active region and lower resolution elsewhere. The twisted rope obtained with XTRAPOL is fully recovered. **a**, Global view showing the disk of tetrahedral-cell mesh and the spherical photosphere, where we have indicated the various resolutions for the

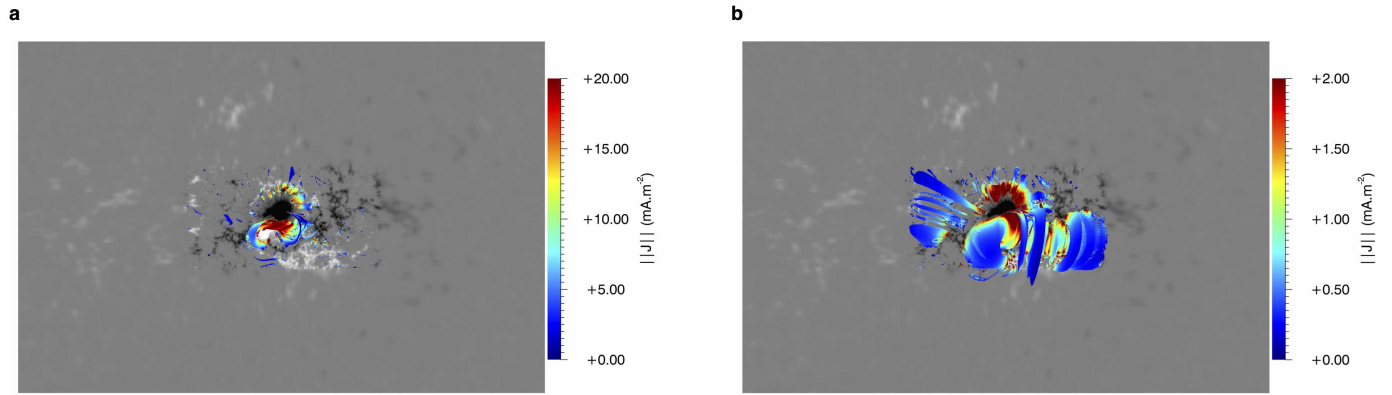
northern hemisphere and, in transparency behind the disk, for the southern hemisphere. **b**, Zoom onto the active region showing the high resolution used around it. **c**, Closer look at the rope, with a cut showing how the adaptive scheme allows high mesh resolution in the regions where the coronal electric current and magnetic field are stronger. **d**, Another point of view, exhibiting the large extent of the rope, which is still confined by the overlying field lines (orange).





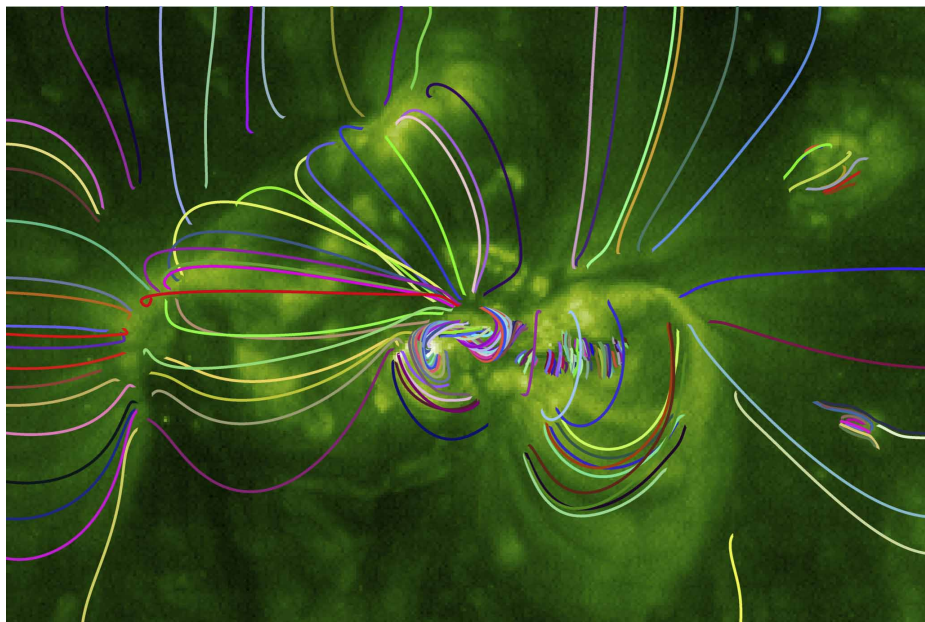
**Extended Data Figure 2 | Hyperbolic flux tube.** **a**, Breaking of the twisted rope into various components to exhibit its hyperbolic nature, using the same colour code as in Fig. 1. **b**, Core of the rope, which is highly twisted (by about  $2.25\pi$ ). **c**, Underlying highly sheared arcs below the core. **d**, Two

J-shaped arcs whose central parts become tangential to each other. **e**, One of the J-shaped set of loops (green) above the sheared arcs (yellow), becoming tangential to each other near the neutral line.



**Extended Data Figure 3 | Signature of the pre-eruptive current density of the reconstructed magnetic configuration of 12 December, 20:30 UT.** We have plotted here two isosurfaces of the force-free function  $\alpha$  measuring the ratio of the electric current density to the magnetic field: **a**,  $\alpha = -0.23 \text{ Mm}^{-1}$ ; **b**,  $\alpha = -0.05 \text{ Mm}^{-1}$ . These isosurfaces are coloured according to the values of

the modulus of the current density,  $|j|$ . The large-scale structure of the twisted rope (and of small parts above) is well exhibited by this quantity  $\alpha$  in **a**, in agreement with Fig. 2b, whereas weaker electric currents (overlying) structures are shown in **b**.



**Extended Data Figure 4 | Extreme-ultraviolet emission and magnetic structure.** Selected field lines of the reconstructed magnetic configuration of 12 December, 20:30 UT, overlaid on an SOHO/EIT extreme-ultraviolet emission image taken at 23:49 UT. The emission is well correlated with the

magnetic lines in the region of the twisted rope and in the regions of approximately current-free loops, such as that located on the right-hand side of the rope.



# Piezoelectricity of single-atomic-layer MoS<sub>2</sub> for energy conversion and piezotronics

Wenzhuo Wu<sup>1\*</sup>, Lei Wang<sup>2\*</sup>, Yilei Li<sup>3</sup>, Fan Zhang<sup>4</sup>, Long Lin<sup>1</sup>, Simiao Niu<sup>1</sup>, Daniel Chenet<sup>4</sup>, Xian Zhang<sup>4</sup>, Yufeng Hao<sup>4</sup>, Tony F. Heinz<sup>3</sup>, James Hone<sup>4</sup> & Zhong Lin Wang<sup>1,5</sup>

The piezoelectric characteristics of nanowires, thin films and bulk crystals have been closely studied for potential applications in sensors, transducers, energy conversion and electronics<sup>1–3</sup>. With their high crystallinity and ability to withstand enormous strain<sup>4–6</sup>, two-dimensional materials are of great interest as high-performance piezoelectric materials. Monolayer MoS<sub>2</sub> is predicted to be strongly piezoelectric, an effect that disappears in the bulk owing to the opposite orientations of adjacent atomic layers<sup>7,8</sup>. Here we report the first experimental study of the piezoelectric properties of two-dimensional MoS<sub>2</sub> and show that cyclic stretching and releasing of thin MoS<sub>2</sub> flakes with an odd number of atomic layers produces oscillating piezoelectric voltage and current outputs, whereas no output is observed for flakes with an even number of layers. A single monolayer flake strained by 0.53% generates a peak output of 15 mV and 20 pA, corresponding to a power density of 2 mW m<sup>–2</sup> and a 5.08% mechanical-to-electrical energy conversion efficiency. In agreement with theoretical predictions, the output increases with decreasing thickness and reverses sign when the strain direction is rotated by 90°. Transport measurements show a strong piezotronic effect in single-layer MoS<sub>2</sub>, but not in bilayer and bulk MoS<sub>2</sub>. The coupling between piezoelectricity and semiconducting properties in two-dimensional nanomaterials may enable the development of applications in powering nanodevices, adaptive bioprobes and tunable/stretchable electronics/optoelectronics.

Crystal structure and symmetry dictate the physical properties of a material and its interaction with external stimuli. Materials with polarization domains, such as Pb(Ti,Zr)O<sub>3</sub>, or with non-centrosymmetric structure, such as ZnO and GaN, are piezoelectric and have wide applications in sensors, transducers, power generation and electronics<sup>1,3,9</sup>. Layered materials, such as graphite, hexagonal boron nitride (h-BN) and many transition-metal dichalcogenides, are centrosymmetric in their bulk three-dimensional form but may exhibit different symmetry when thinned down to a single atomic layer<sup>7,10,11</sup>. In graphene, inversion symmetry is preserved because both atoms in the unit cell are identical, whereas monolayer h-BN and transition-metal dichalcogenides become non-centrosymmetric because of the absence of an inversion centre, which leads to novel properties such as valley-selective circular dichroism<sup>12,13</sup> and large second-order nonlinear susceptibility<sup>14,15</sup>, corresponding to the optical second-harmonic generation (SHG) process<sup>16</sup>. Single-atomic-layer h-BN, MoS<sub>2</sub>, MoSe<sub>2</sub> and WTe<sub>2</sub> have also been theoretically predicted to show piezoelectricity as a result of strain-induced lattice distortion and the associated ion charge polarization<sup>7,8,11</sup>, suggesting possible applications of two-dimensional (2D) nanomaterials in nanoscale electromechanical devices that take advantage of their outstanding semiconducting and mechanical properties<sup>4–6,17</sup>. Here we report an experimental observation of piezoelectricity in single-atomic-layer 2D MoS<sub>2</sub> and its application in mechanical energy harvesting and piezotronic sensing. Cyclic stretching and releasing of odd-layer MoS<sub>2</sub> flakes produces oscillating electrical outputs, which converts mechanical energy into electricity.

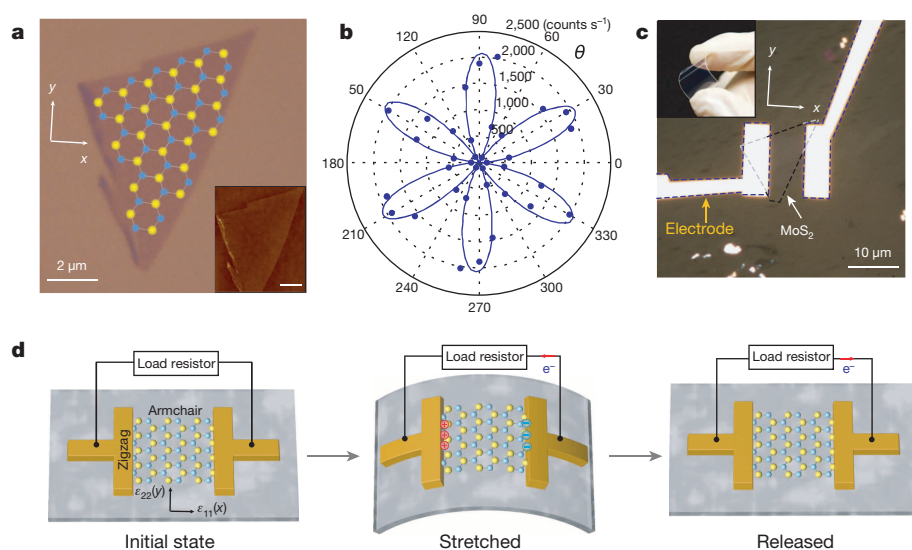
The strain-induced polarization charges in single-layer MoS<sub>2</sub> can also modulate charge carrier transport at the MoS<sub>2</sub>–metal barrier and enable enhanced strain sensing. In addition, we have also observed large piezoresistivity in even-layer MoS<sub>2</sub> with a gauge factor of about 230 for the bilayer material, which indicates a possible strain-induced change in band structure<sup>18</sup>. Our study demonstrates the potential of 2D nanomaterials in powering nanodevices, adaptive bioprobes and tunable/stretchable electronics/optoelectronics.

In our experiments, MoS<sub>2</sub> flakes were mechanically exfoliated onto a polymer stack consisting of water-soluble polyvinyl alcohol and poly(methyl methacrylate) on a Si substrate, with the total polymer thickness tuned to be 275 nm for good optical contrast. Few-layer MoS<sub>2</sub> flakes were first identified under an optical microscope. The layer thickness was then measured by atomic force microscopy (Fig. 1a, inset) and confirmed by Raman spectroscopy (Extended Data Fig. 1a). SHG was subsequently used to determine the crystallographic orientation of the MoS<sub>2</sub> flakes (Methods and Extended Data Fig. 1b). Figure 1b shows a polar plot of the second-harmonic (SH) signal intensity from a single-layer MoS<sub>2</sub> flake as a function of the crystal's azimuthal angle. Here we measured the SH component perpendicular to the excitation polarization. The lattice orientation was determined by fitting the angle dependence of SH intensity with  $I = I_0 \sin^2(3\theta)$ , where  $\theta$  denotes the angle between the direction of the 'armchair' edge and the polarization of the excitation laser, and  $I_0$  is the maximum intensity of the SH response. Figure 1a schematically depicts the derived lattice orientation superimposed on the optical image of a single-layer MoS<sub>2</sub> flake; the  $x$  axis is taken to be along the 'armchair' direction, and the  $y$  axis along the 'zigzag' direction. After optical characterization, flakes were subsequently transferred to a polyethylene terephthalate (PET) flexible substrate using methods described previously<sup>19</sup>. Electrical contacts made of Cr/Pd/Au (1 nm/20 nm/50 nm) were deposited with the metal–MoS<sub>2</sub> interface parallel to the  $y$  axis. Figure 1c shows a typical flexible device with the single-layer MoS<sub>2</sub> flake outlined by black dashed line. When the substrate was bent mechanically, uniaxial strain is applied to the MoS<sub>2</sub> with a magnitude proportional to the inverse bending radius (Fig. 1d and Methods). The applied strain was limited to 0.8% to avoid sample slippage<sup>20</sup>. We studied the piezoelectric response by applying strain to a device coupled to an external load resistor (Fig. 1d). In this configuration, strain-induced polarization charges at the sample edges can drive the flow of electrons in external circuit<sup>9</sup>. When the substrate is released, electrons flow back in the opposite direction.

Figure 2a and Extended Data Fig. 3 show the piezoelectric current and voltage responses of a single-layer MoS<sub>2</sub>. When strain is applied in the  $x$  direction ('armchair' direction), positive voltage and current output were observed with increasing strain, and negative output was observed with decreasing strain, directly demonstrating the conversion of mechanical energy into electricity<sup>9</sup> (see also Extended Data Fig. 4 and Methods). Both responses increased with the magnitude of the applied

<sup>1</sup>School of Materials Science and Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0245, USA. <sup>2</sup>Department of Electrical Engineering, Columbia University, New York, New York 10027, USA. <sup>3</sup>Department of Physics, Columbia University, New York, New York 10027, USA. <sup>4</sup>Department of Mechanical Engineering, Columbia University, New York, New York 10027, USA. <sup>5</sup>Beijing Institute of Nanoenergy and Nanosystems, Chinese Academy of Sciences, 100083 Beijing, China.

\*These authors contributed equally to this work.



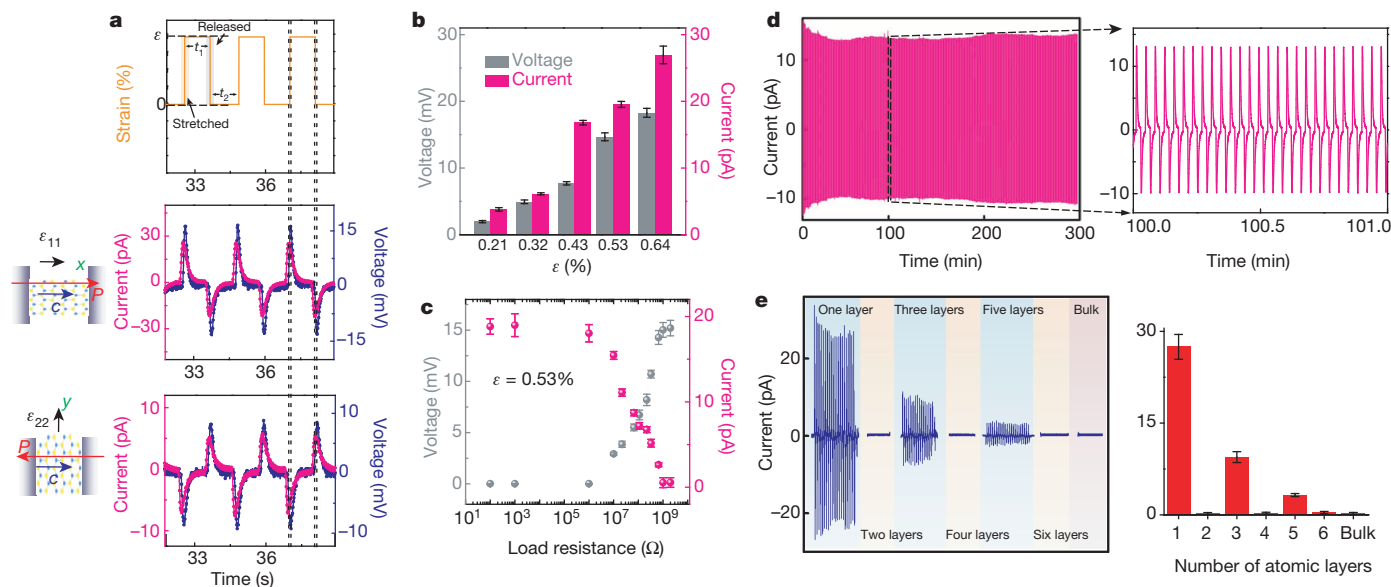
**Figure 1 | Single-layer MoS<sub>2</sub> piezoelectric device and operation scheme.** **a**, Optical image of the single-atomic layer MoS<sub>2</sub> flake with superimposed lattice orientation derived from SHG results. Blue and yellow spheres represent Mo and S atoms, respectively. Inset: atomic force microscopy image of the flake. Scale bar 2 μm. **b**, Polar plot of the SH intensity from single-layer MoS<sub>2</sub> as a function of the crystal's azimuthal angle  $\theta$ . The symbols are experimental data and the solid lines are fits to the symmetry analysis described in the text. **c**, A typical flexible device with single-layer MoS<sub>2</sub> flake and electrodes at its zigzag edges. Inset: optical image of the flexible device. **d**, Operation scheme of the single-layer MoS<sub>2</sub> piezoelectric device. When the device is stretched, piezoelectric polarization charges of opposite polarity (plus and minus symbols) are induced at the zigzag edges of the MoS<sub>2</sub> flake. Periodic stretching and releasing of the substrate can generate piezoelectric outputs in external circuits with alternating polarity (as indicated by the red arrows).

strain: for single-layer MoS<sub>2</sub> with width of  $\sim 5 \mu\text{m}$  and a length of  $\sim 10 \mu\text{m}$ , the peak open-circuit voltage reached 18 mV and the peak short-circuit current reached 27 pA (Fig. 2b), with voltage and current responsivities of  $55.1 \pm 12.3 \text{ pA}$  and  $32.8 \pm 4.5 \text{ mV}$ , respectively, for each 1% change in strain. There were no significant electrical outputs from bare PET substrates without a single-layer MoS<sub>2</sub> flake (Extended Data Fig. 5b).

The dependence of piezoelectric charge polarization on the directions of principal strain in 2D materials was also investigated. The coupling between polarization ( $P_i$ ) and strain ( $\epsilon_{jk}$ ) tensor can be quantified to first order by the third-rank piezoelectric tensor  $e_{ijk} = (\partial P_i / \partial \epsilon_{jk})$ , where  $i, j, k \in (1, 2, 3)$ , with 1, 2 and 3 corresponding to the  $x$ ,  $y$  and  $z$

axes, respectively. Symmetry analysis of the  $D_{3h}$  point group suggested that there was only one non-zero independent coefficient  $e_{11}$  for single-layer MoS<sub>2</sub>. The in-plane polarization along the  $x$  axis, sensed by the metal electrodes as shown in Fig. 1d, can be expressed as  $P_1 = e_{11}(\epsilon_{11} - \epsilon_{22})$ , whereas  $P_2$  along the  $y$  axis is related to the pure shear strain  $\epsilon_{12}$  and can be ignored in these experiments. A distinctive consequence of this symmetry is that the output is expected to reverse sign when the strain is rotated from the  $x$  ('armchair') to the  $y$  ('zigzag') direction. This was verified experimentally, as shown in the bottom panel of Fig. 2a.

To quantify the power output of the piezoelectric circuit, it is necessary to study the voltage and current outputs as a function of load resistance, as shown in Fig. 2c (see Extended Data Fig. 6 for circuit details). The



**Figure 2 | Piezoelectric outputs from single-layer and multi-layer MoS<sub>2</sub> devices.** **a**, Voltage response with 1 GΩ external load and short-circuit current response of a single-layer MoS<sub>2</sub> device under periodic strain in two different principal directions. Top: applied strain as a function of time. Middle: corresponding piezoelectric outputs from single-atomic-layer MoS<sub>2</sub> when strain is applied in the  $x$  direction (armchair direction). Bottom: corresponding piezoelectric outputs from the same device when strain is applied in the  $y$  direction (zigzag direction). The phase difference highlighted by black dashed lines is obtained by theoretical derivation and has been intentionally exaggerated for clarity, not experimental measurement. Red, blue and black arrows represent the directions of polarization, the polar axis of MoS<sub>2</sub> and

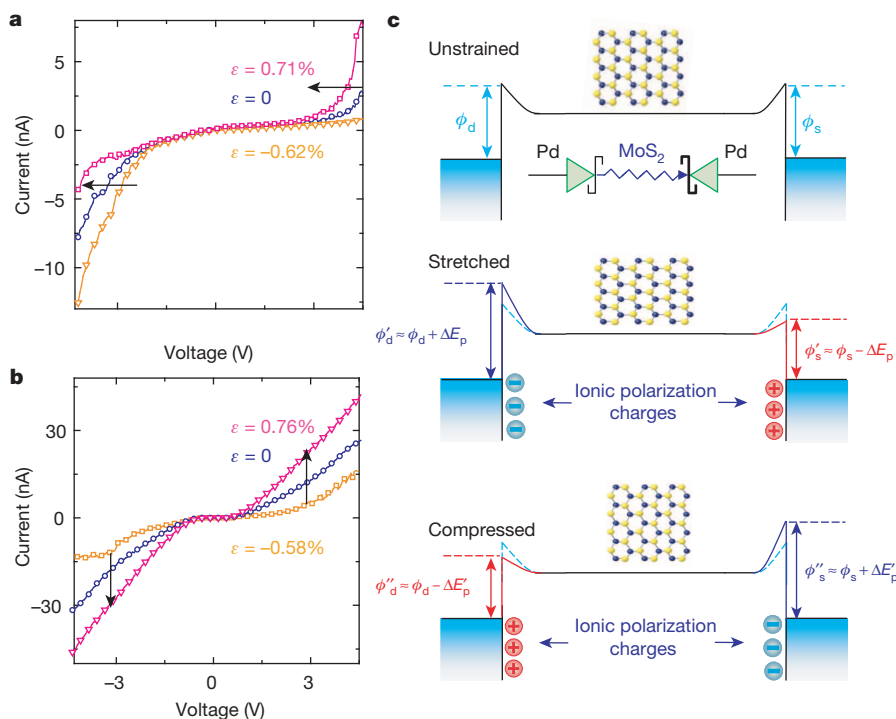
principal strains, respectively. **b**, Dependence of piezoelectric outputs from a single-layer MoS<sub>2</sub> device on the magnitude of the applied strain. Mean values from 20 technical replicates are indicated. Error bars represent s.d. **c**, Dependence of voltage and current outputs from a single-layer MoS<sub>2</sub> device under 0.53% strain as a function of load resistance. Mean values from 20 technical replicates are indicated. Error bars represent s.d. **d**, Cyclic test showing the stability of single-layer MoS<sub>2</sub> device for prolonged period. **e**, Evolution of the piezoelectric outputs with increasing number of atomic layers ( $n$ ) in MoS<sub>2</sub> flakes. For each device, mean values from 20 technical replicates are indicated. Error bars represent s.d.

output current was constant for a load resistance of up to  $\sim 1\text{ M}\Omega$  and then decreased with increasing load, whereas the output voltage was  $\sim 0\text{ V}$  and began to increase at the same point. The maximum instantaneous power delivered to the load at 0.53% strain was achieved for a load resistance of  $\sim 220\text{ M}\Omega$  and reached 55.3 fW ( $5.53 \times 10^{-14}\text{ W}$ ), with a corresponding power density of  $2\text{ mW m}^{-2}$  (Extended Data Fig. 6). The conversion efficiency of the single-layer MoS<sub>2</sub> nanogenerator, which is the ratio of the electric power delivered to the load to the total mechanical deformation energy stored in the single-layer MoS<sub>2</sub> after being strained, can therefore be estimated as  $\sim 5.08\%$  (Methods). This energy conversion was stable over time, as shown for cyclic loading up to 0.43% strain at 0.5 Hz for 300 min (Fig. 2d and Extended Data Fig. 7). The observed slight decrease in output may have been caused by mechanical fatigue of the flexible substrate<sup>9</sup>.

We next examined the evolution of the piezoelectric signal with an increasing number of atomic layers ( $n$ ). As discussed above, because of the opposite orientation of alternating layers in the most common (2H) form of MoS<sub>2</sub>, flakes with  $n = \text{even}$  are expected to be centrosymmetric and thus lose both their piezo response and SHG signal<sup>18,16</sup>. For the same reason, in samples with an odd number of layers, the piezo response and SHG should return. Figure 2e shows the measured piezoelectric output for MoS<sub>2</sub> flakes with  $n = 1, 2, 3, 4, 5$  and 6, and for a bulk MoS<sub>2</sub> flake with a thickness of more than 100 nm. Consistent with the above picture, the SHG intensity was strong for  $n = 1, 3, 5$  and disappeared for  $n = 2, 4, 6$  and the bulk (Extended Data Fig. 8), consistent with previous reports<sup>16</sup>. To measure the piezoelectric response, the source and drain electrodes were made at the zigzag edge for the odd-layer flakes and at an arbitrary angle for the even-layer flakes, because of the absence of a SHG signal. Almost no detectable output can be seen for bulk flake and even-layer samples. For odd-layer samples the piezoelectric output is large and decreases roughly as the inverse of  $n$ . These results confirm that single-layer MoS<sub>2</sub> with broken inversion symmetry has a strong intrinsic piezoelectric response, whereas centrosymmetric bilayers and bulk crystals are non-piezoelectric<sup>7,8</sup>.

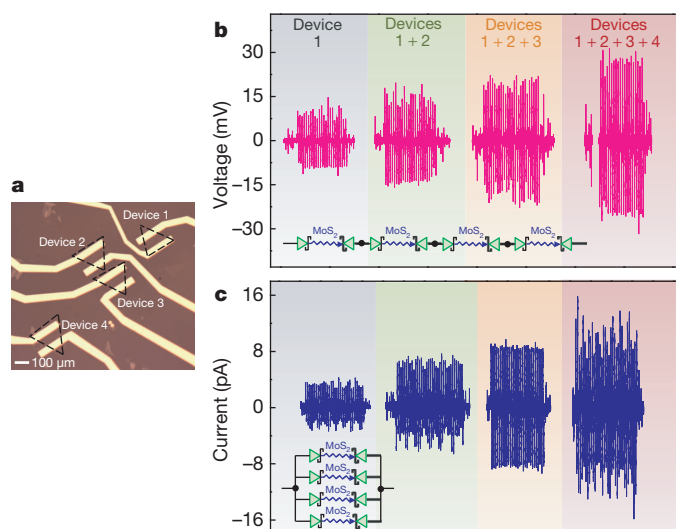
We next characterized the changes in direct-current electrical transport properties of the devices with strain, in a two-terminal configuration with the polarity of the applied voltage defined with respect to the drain electrode. The metal–semiconductor–metal (MSM) device as fabricated consisted of two back-to-back Schottky barriers, and transport

across the reverse-biased Pd–MoS<sub>2</sub> Schottky barrier limited the current flow<sup>21</sup>. In this configuration, changes in transport behaviour arose largely from two effects: the piezotronic effect<sup>22</sup>, in which strain-induced charge asymmetrically modulated the Schottky barriers, and a piezoresistive effect, in which strain-induced bandgap change modulated the entire resistance of the device. For the single-layer device, the current–voltage curve shifted leftwards (towards negative drain bias) under tensile strain, and rightwards with compressive strain (Fig. 3a). The opposite trend was observed under negative drain bias. This asymmetric modulation was similar to the piezotronic effect reported for piezoelectric semiconductors with wurtzite and zincblende structures, in which the modulation of carrier transport arises as a result of piezoelectric polarization in the crystal (Methods). Here, piezoelectric polarization charges at the zigzag edges were able to affect the metal–MoS<sub>2</sub> contacts directly (Fig. 3c) by modifying the concentration or distribution of free carriers in MoS<sub>2</sub> as well as by modulating the electronic charges in interface states<sup>22</sup>, such that the mechanical strain functioned as a controlling gate signal. Two points should be noted here. First, MoS<sub>2</sub> contacted Pd electrodes in two ways in our experiment: at its zigzag end edges, where the piezoelectric polarization charges were distributed, and at the top surface, which may not have had piezoelectric charges. Although the polarization charges were induced at the zigzag edge interface, they were still able to affect the electrical transport across the whole Schottky contact area formed between MoS<sub>2</sub> and metal. This was probably due to the fact that the majority electric field under bias was distributed at the end-edge contacts, in an analogous manner to previous studies on piezotronic effect in ZnO nanowires, in which the electric field was focused at the end surface although the contact also occurred at the side surfaces<sup>23</sup>. Second, the free charge carriers in monolayer MoS<sub>2</sub> were not taken into consideration for simplicity in the band diagrams (Fig. 3c). In reality, the finite carrier density in MoS<sub>2</sub> (which depended on factors such as intrinsic and environmental doping and varied from flake to flake) may have resulted in partial screening of the strain-induced polarization charges and hence could have affected its piezoelectric performance (Methods), but it is still possible to observe its piezoelectric power output and piezotronic transport characteristics. The results in Fig. 3a can be used to determine the crystallographic orientation of the flake uniquely as having the S-edge and Mo-edge at the drain and source electrodes, respectively. Controllable modulation of M–S contacts or  $p$ – $n$  junctions in 2D nanomaterials by



**Figure 3 | Direct-current electrical characterizations of single-layer and bilayer MoS<sub>2</sub> devices under strains.** **a**, The asymmetric modulation of carrier transport by strains under opposite drain bias in a single-layer device shows characteristics of a piezotronic effect. **b**, The symmetrical modulation of carrier transport by strains under opposite drain bias in a bilayer device shows characteristics of a piezoresistive effect. **c**, Band diagrams explaining the piezotronic behaviour observed in a single-layer device as a result of the changes in Schottky barrier heights by strain-induced polarization.  $\phi_d$  and  $\phi_s$  represent the Schottky barrier heights formed at drain and source contacts, respectively.  $E_p$  indicates the change in Schottky barrier height by piezoelectric polarization charges.





**Figure 4** | Array integration of CVD single-layer MoS<sub>2</sub> flakes. **a**, Optical image of an array consisting of four CVD single-layer MoS<sub>2</sub> flakes. **b**, Constructive voltage outputs by serial connection of the individual flakes in the circuit. **c**, Constructive current outputs by parallel connection of the individual flakes in the circuit.

strain-induced polarization may offer a novel approach unavailable to conventional technologies using electrical control signals, without modifying the interface structure or chemistry, for implementing tunable electronics/optoelectronics, enhanced photovoltaics, hybrid spintronics and catalysis<sup>24,25</sup>.

In bilayer and bulk MoS<sub>2</sub> devices, the response is purely piezoresistive: the current increases symmetrically with applied strain, and the gap region in the  $I$ - $V$  curve shrinks, consistent with a lowering of both source and drain Schottky barriers as a result of a decrease in the bandgap<sup>18</sup> and/or a change in carrier density (see Fig. 3b for a bilayer device and Extended Data Fig. 9 for a bulk device). The gauge factors ( $[\Delta I(\epsilon)]/I(0)/\Delta\epsilon$ ) of the bilayer ( $\sim 230$ ) and bulk MoS<sub>2</sub> devices ( $\sim 200$ ) measured in our experiments were comparable to that reported for a state-of-the-art silicon strain sensor ( $\sim 200$ ), and exceeded the values of conventional metal strain gauges ( $\sim 1$ – $5$ ) and a graphene strain sensor ( $\sim 2$ ) (ref. 26). This is a preliminary demonstration of large piezoresistivity in 2D MoS<sub>2</sub>, and motivates further study for using it in highly sensitive strain sensing.

Finally, we demonstrate the array integration of single-layer MoS<sub>2</sub> flakes to boost the piezoelectric output for energy conversion (Fig. 4). High-quality single-layer MoS<sub>2</sub> crystals were grown by seed-free chemical vapour deposition (CVD) on a Si/SiO<sub>2</sub> substrate, using a method reported previously<sup>27</sup>. Single-domain triangles were then transferred from the growth substrate to flexible PET<sup>19</sup>. Previous studies noted that molybdenum zigzag and sulphur zigzag are the two dominant morphologies of CVD MoS<sub>2</sub> triangles<sup>28</sup>, and molybdenum zigzag triangles consistently have sharper and straighter edges than sulphur zigzag triangles<sup>27</sup>. This morphological difference allowed us to easily identify the crystal orientation of CVD MoS<sub>2</sub> in optical images and pattern the electrodes accordingly (Fig. 4a). For this demonstration, four CVD MoS<sub>2</sub> flakes were chosen. By constructively connecting the four flakes either in series or in parallel, consistent enhancements in output voltages or currents were observed (Fig. 4b, c). Moreover, by destructively connecting devices 1 and 2 either in series or in parallel, the combined outputs were the difference of the two individual outputs (Extended Data Fig. 10). This may open up possibilities of achieving practical technology at an even larger scale with 2D piezoelectric nanomaterials for powering nanodevices, tactile imaging and wearable electronics<sup>9,29</sup>. Nevertheless, efforts are required to achieve a better understanding of the synthesis of CVD MoS<sub>2</sub> with more controllable properties and optimize large-scale fabrication for improving device-to-device uniformity and achieving practical applications based on an array of single-atomic-layer MoS<sub>2</sub> devices.

The piezoelectricity and large mechanical stretchability and flexibility of single-atomic-layer MoS<sub>2</sub> demonstrate its potential applications in electromechanical sensing, wearable technology, pervasive computing and implanted devices. The integration of a MoS<sub>2</sub>-based power source with graphene and other functional units or devices based on 2D materials for energy storage, sensing, logic computation and communication on the same substrate may permit the construction of an atomic-thin self-powered nanosystem that can operate self-sustainably without external bias by harvesting energy from the ambient environment<sup>30</sup>, especially in circumstances in which other energy sources, such as solar or thermal energy, are not readily available.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 23 May; accepted 8 August 2014.

Published online 15 October 2014.

- Kingon, A. I. & Srinivasan, S. Lead zirconate titanate thin films directly on copper electrodes for ferroelectric, dielectric and piezoelectric applications. *Nature Mater.* **4**, 233–237 (2005).
- Ferren, R. A. Advances in polymeric piezoelectric transducers. *Nature* **350**, 26–27 (1991).
- Wang, Z. L. & Song, J. H. Piezoelectric nanogenerators based on zinc oxide nanowire arrays. *Science* **312**, 242–246 (2006).
- Lee, C., Wei, X. D., Kysar, J. W. & Hone, J. Measurement of the elastic properties and intrinsic strength of monolayer graphene. *Science* **321**, 385–388 (2008).
- Lee, G. H. et al. High-strength chemical-vapor deposited graphene and grain boundaries. *Science* **340**, 1073–1076 (2013).
- Bertolazzi, S., Brivio, J. & Kis, A. Stretching and breaking of ultrathin MoS<sub>2</sub>. *ACS Nano* **5**, 9703–9709 (2011).
- Duerloo, K. A. N., Ong, M. T. & Reed, E. J. Intrinsic piezoelectricity in two-dimensional materials. *J. Phys. Chem. Lett.* **3**, 2871–2876 (2012).
- Michel, K. H. & Verberck, B. Phonon dispersions and piezoelectricity in bulk and multilayers of hexagonal boron nitride. *Phys. Rev. B* **83**, 115328 (2011).
- Yang, R. S., Qin, Y., Dai, L. M. & Wang, Z. L. Power generation with laterally packaged piezoelectric fine wires. *Nature Nanotechnol.* **4**, 34–39 (2009).
- Ong, M. T. & Reed, E. J. Engineered piezoelectricity in graphene. *ACS Nano* **6**, 1387–1394 (2012).
- Michel, K. H. & Verberck, B. Theory of elastic and piezoelectric effects in two-dimensional hexagonal boron nitride. *Phys. Rev. B* **80**, 224301 (2009).
- Mak, K. F., He, K. L., Shan, J. & Heinz, T. F. Control of valley polarization in monolayer MoS<sub>2</sub> by optical helicity. *Nature Nanotechnol.* **7**, 494–498 (2012).
- Zeng, H. L., Dai, J. F., Yao, W., Xiao, D. & Cui, X. D. Valley polarization in MoS<sub>2</sub> monolayers by optical pumping. *Nature Nanotechnol.* **7**, 490–493 (2012).
- Kumar, N. et al. Second harmonic microscopy of monolayer MoS<sub>2</sub>. *Phys. Rev. B* **87**, 161403 (2013).
- Malard, L. M., Alencar, T. V., Barboza, A. P. M., Mak, K. F. & de Paula, A. M. Observation of intense second harmonic generation from MoS<sub>2</sub> atomic crystals. *Phys. Rev. B* **87**, 201401 (2013).
- Li, Y. L. et al. Probing symmetry properties of few-layer MoS<sub>2</sub> and h-BN by optical second-harmonic generation. *Nano Lett.* **13**, 3329–3333 (2013).
- Wang, Q. H., Kalantar-Zadeh, K., Kis, A., Coleman, J. N. & Strano, M. S. Electronics and optoelectronics of two-dimensional transition metal dichalcogenides. *Nature Nanotechnol.* **7**, 699–712 (2012).
- Conley, H. J. et al. Bandgap engineering of strained monolayer and bilayer MoS<sub>2</sub>. *Nano Lett.* **13**, 3626–3630 (2013).
- Dean, C. R. et al. Boron nitride substrates for high-quality graphene electronics. *Nature Nanotechnol.* **5**, 722–726 (2010).
- He, K., Poole, C., Mak, K. F. & Shan, J. Experimental demonstration of continuous electronic structure tuning via strain in atomically thin MoS<sub>2</sub>. *Nano Lett.* **13**, 2931–2936 (2013).
- Neal, A. T., Liu, H., Gu, J. J. & Ye, P. D. in *Conference Digest, Device Research Conference (DRC), 18–20 June 2012 70th Annual* 65–66, <http://dx.doi.org/10.1109/DRC.2012.6256928> (IEEE, 2012).
- Wang, Z. L. Piezopotential gated nanowire devices: piezotronics and piezophotonics. *Nano Today* **5**, 540–552 (2010).
- Hu, Y. F. et al. Temperature Dependence of the Piezotronic Effect in ZnO Nanowires. *Nano Lett.* **13**, 5026–5032 (2013).
- Chen, J. R. et al. Control of Schottky barriers in single layer MoS<sub>2</sub> transistors with ferromagnetic contacts. *Nano Lett.* **13**, 3106–3110 (2013).
- Yang, H. et al. Graphene barristor, a triode device with a gate-controlled Schottky barrier. *Science* **336**, 1140–1143 (2012).
- Huang, M. Y., Pascal, T. A., Kim, H., Goddard, W. A. & Greer, J. R. Electronic-mechanical coupling in graphene from in situ nanoindentation experiments and multiscale atomistic simulations. *Nano Lett.* **11**, 1241–1246 (2011).
- van der Zande, A. M. et al. Grains and grain boundaries in highly crystalline monolayer molybdenum disulphide. *Nature Mater.* **12**, 554–561 (2013).

28. Lauritsen, J. V. *et al.* Size-dependent structure of MoS<sub>2</sub> nanocrystals. *Nature Nanotechnol.* **2**, 53–58 (2007).
29. Wu, W. Z., Wen, X. N. & Wang, Z. L. Taxel-addressable matrix of vertical-nanowire piezotronic transistors for active and adaptive tactile imaging. *Science* **340**, 952–957 (2013).
30. Wang, Z. L. & Wu, W. Z. Nanotechnology-enabled energy harvesting for self-powered micro-/nanosystems. *Angew. Chem. Int. Ed.* **51**, 11700–11721 (2012).

**Acknowledgements** We thank C. Dean, P. Kim and K. Shepard for discussions, and K. Alexander and E. Reed for theoretical consultation. This research was supported by the US Department of Energy, Office of Basic Energy Sciences (DE-FG02-07ER46394) and US National Science Foundation (DMR-1122594). Z.L.W. acknowledges the 'Thousands Talents' programme for pioneer researcher and his innovation team, National Natural Science Foundation of China (51432005), and China and

Beijing City Committee of Science and Technology (Z131100006013004 and Z131100006013005).

**Author Contributions** W.Z.W., L.W., Z.L.W. and J.H. conceived the idea. W.Z.W., L.W., Y.L.L., T.F.H., J.H. and Z.L.W. designed the experiments. L.W., D.C., F.Z. and Y.F.H. synthesized the material. L.W., D.C. and W.Z.W. fabricated the devices. Y.L.L. performed the SHG characterization. W.Z.W., L.W., L.L., Y.L.L. and S.M.N. conducted the experiments. W.Z.W., L.W., J.H. and Z.L.W. analysed the data and wrote the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Z.L.W. ([zhong.wang@mse.gatech.edu](mailto:zhong.wang@mse.gatech.edu)) or J.H. ([jh2228@columbia.edu](mailto:jh2228@columbia.edu)).

## METHODS

### Determination of crystallographic orientation in MoS<sub>2</sub> flakes by optical SHG.

The SHG measurements were performed in reflection geometry (Extended Data Fig. 1b). The incident excitation beam was normal to the sample. The pump radiation was supplied by a mode-locked Ti:sapphire oscillator operating at a repetition rate of 80 MHz. The pulses were of 90 fs duration and centred on a wavelength of 810 nm. Using a  $\times 100$  objective, we focused the pump beam on the sample with a spot size of about 1  $\mu\text{m}$ . We limited the average laser power to 1 mW. The retro-reflected SH signal was collected by the same objective, separated by a beam splitter and filtered by a short-pass optical filter (cutoff at 785 nm) to block the reflected fundamental radiation. An analyser was used to select the polarization component of the SH radiation perpendicular to the polarization of the pump beam. After dispersal in a spectrometer, the SH signal was detected by a liquid-nitrogen-cooled charge-coupled device camera. The SH character of the detected radiation was verified by its wavelength and quadratic power dependence on the pump intensity. In our setup we could freely rotate the samples to obtain the orientation dependence of the SH response. However, it should be noted that the lattice orientation cannot be uniquely determined by SHG because the SH signal remains the same if the lattice plane is rotated by  $180^\circ$  with respect to its normal direction. In other words, only the direction of the Mo–S bond axis can be obtained from the SHG signal, and no conclusion can be drawn about which side is Mo-edge or S-edge.

**Electrical output measurements.** A programmable electrometer (part number 6514, Keithley) with 200 T $\Omega$  input impedance was used for measuring voltage signals from the device. A low-noise current preamplifier (part number SR570; Stanford Research Systems, Inc.) was used for current measurements, with its direct-current input impedance at 1 M $\Omega$  for the sensitivity level used ( $10^{-11}$  and  $10^{-12}$  A V $^{-1}$ ). A 10-Hz low-pass filter was used for both voltage and current measurements. A computer-controlled measurement software written in Labview was used to collect and record the data. All measurements were made inside a home-built electrical cage. A linear motor (LinMot PS01-23  $\times$  80) was used for applying programmed driving strain inputs.

**Estimation of strain induced in MoS<sub>2</sub> device.** Because the dimensions of MoS<sub>2</sub> flakes ( $5 \mu\text{m} \times 5 \mu\text{m} \times (0.6\text{--}100 \text{ nm})$ ) are significantly smaller than those of the PET substrates ( $2.5 \text{ cm} \times 2 \text{ cm} \times 500 \mu\text{m}$ ), the mechanical behaviour of the substrate and entire device is not affected by the MoS<sub>2</sub> flake. Consequently, the Saint-Venant theory for small bending deformation can be adopted for estimating the induced strain in MoS<sub>2</sub> devices<sup>31</sup>. The PET substrate can be approximated as a beam structure of thickness  $a$ , width  $w$  and length  $l$ . The origin for calculation is defined as the centre of the fixed edge. The  $x$  and  $z$  axis are along the width ( $w$ ) and length ( $l$ ) directions of the substrate, respectively (Extended Data Fig. 2a). Because the MoS<sub>2</sub> flake lies above the neutral strain axis of the entire device, the deflection of the substrate under an external force  $f_y$ , exerted by the linear motor, results in a pure elongation or contraction in the MoS<sub>2</sub> flake if no slippage is considered. It follows that the uniaxial in-plane strain induced in MoS<sub>2</sub> can be estimated by the  $\varepsilon_{zz}$  component of strain in the PET substrate. Applying the Saint-Venant theory for small deflections of the beam,  $\sigma_{xz} = -(f_y/I_{xx})y(l-z)$ ,  $\sigma_{xx} = \sigma_{yy} = 0$ , in which  $I_{xx}$  is the moment of inertia for the beam. Consequently,  $\varepsilon_{zz} = \sigma_{zz}/E$ , where  $E$  is the Young's modulus of substrate. Experimentally, it is more convenient to measure the lateral deflection  $D_{\text{max}}$  of the free end of substrate rather than the external point force  $f_y$ . Since  $D_{\text{max}} = f_y l^3 / 3EI_{xx}$  by classical beam mechanics, then

$$\varepsilon_{zz} = -\frac{3y D_{\text{max}}}{l} \left(1 - \frac{z}{l}\right)$$

Because the MoS<sub>2</sub> flake is lying at  $y = \pm a/2$  and  $z = z_0$ , where  $z_0$  is the distance between the fixed edge and the MoS<sub>2</sub> flake, which can be measured in the experiment, the strain induced in MoS<sub>2</sub> flake can be estimated by

$$\varepsilon = \varepsilon_{zz} = \mp \frac{3}{2} \frac{a D_{\text{max}}}{l} \left(1 - \frac{z_0}{l}\right) \quad (1)$$

The negative sign is for compressive strain and positive sign is for tensile strain. All variables in the above equation can be readily measured experimentally. In the experiment, the linear motor used for inducing strain is first accelerated and then decelerated with a constant acceleration  $a$ . The maximum lateral deflection  $D_{\text{max}}$  of the free end of the substrate, the acceleration  $a$  and the hold time ( $t_1$  and  $t_2$ ) are known parameters and can be controlled by the linear motor (see Extended Data Fig. 2b, c). Therefore the driving signal (moving distance  $d$ ) in one cycle can be mathematically described by the following equations:

$$d = at^2/2, (t < \sqrt{D_{\text{max}}/a})$$

$$d = D_{\text{max}} - \frac{1}{2}a \left(2\sqrt{D_{\text{max}}/a} - t\right)^2, (\sqrt{D_{\text{max}}/a} \leq t < 2\sqrt{D_{\text{max}}/a})$$

$$d = D_{\text{max}}, (2\sqrt{D_{\text{max}}/a} \leq t < 2\sqrt{D_{\text{max}}/a} + t_1)$$

$$d = D_{\text{max}} - \frac{1}{2}a \left(t - 2\sqrt{D_{\text{max}}/a} - t_1\right)^2, (2\sqrt{D_{\text{max}}/a} + t_1 \leq t < 3\sqrt{D_{\text{max}}/a} + t_1)$$

$$d = \frac{1}{2}a \left(t - 4\sqrt{D_{\text{max}}/a} - t_1\right)^2, (3\sqrt{D_{\text{max}}/a} + t_1 \leq t < 4\sqrt{D_{\text{max}}/a} + t_1)$$

$$d = 0, (4\sqrt{D_{\text{max}}/a} + t_1 \leq t < 4\sqrt{D_{\text{max}}/a} + t_1 + t_2)$$

A representative plot is shown in Fig. 2a and Extended Data Fig. 2b. As a result of the large difference between scales of  $a$  ( $5\text{--}10 \text{ m s}^{-2}$ ) and  $D_{\text{max}}$  ( $< 10 \text{ mm}$ ), the stretch and release edges of the strain curve are very sharp. Snapshots from the typical configurations of the linear motor are also included in Extended Data Fig. 2c.

### Power generation by piezoelectric polarization charges in single-layer MoS<sub>2</sub>.

When single-layer MoS<sub>2</sub> is subjected to tensile strain, effective piezoelectric charges are induced at the armchair edges as a result of the polarization of atoms in the strained crystal (Extended Data Fig. 4b). The negative polarization charges deplete the barrier interface and drive the electron flow from the left electrode to the right electrode through an external load (Extended Data Fig. 4b), giving rise to the current peak labelled 'b' in Extended Data Fig. 4. The resistance of the Schottky barrier is significantly high for voltages below a threshold value and thus blocks the flow of electrons through the wire (Extended Data Fig. 4b). The electrons accumulate at the interfacial region between the right electrode and the MoS<sub>2</sub>; the effect of piezoelectric polarization charges is balanced by the accumulated electrons and the Fermi levels in the entire system reach a new equilibrium (Extended Data Fig. 4c). When the tensile strain in the MoS<sub>2</sub> is released, the piezoelectric polarization charges vanish immediately and the electrons previously accumulated at the right electrode flow back to the left electrode through the external load to return the system to the original state, resulting in the downward current peak labelled 'd' in Extended Data Fig. 4. The above process therefore performs one cycle of energy harvesting and conversion from the mechanical to the electrical domain by single-layer MoS<sub>2</sub>.

The piezoelectric voltage constant  $g$  is the electric field generated by a piezoelectric material per unit of mechanical stress applied or, alternatively, is the mechanical strain experienced by a piezoelectric material per unit of electric displacement applied. The first subscript to  $g$  indicates the direction of the electric field generated in the material, or the direction of the applied electric displacement. The second subscript is the direction of the applied stress or the induced strain, respectively. The strength of the induced electric field produced by a piezoelectric material in response to an applied physical stress is the product of the value for the applied stress and the value for  $g$ . For simplicity we can assume that the piezoelectric material is an insulator with no internal screening of the strain-induced polarization charges and that the Schottky barrier between the piezoelectric material and the metal electrode can fully block the electron injection from metal electrode back to the piezoelectric material. Therefore, if we apply a strain  $\varepsilon$  to the piezoelectric material, the internal piezoelectric field generated inside the piezoelectric material is given by  $E = g\varepsilon Y$ , where  $g$  is piezoelectric voltage constant and  $Y$  is the Young's modulus of the piezoelectric material. Thus, the internal piezoelectric potential across the piezoelectric material can be given by  $V_{\text{piezo}} = Eh = g\varepsilon Yh$ , where  $h$  is the length of the material. Thus, at open-circuit condition, the voltage between the two electrodes is this internal piezoelectric potential. Therefore  $V_{\text{OC}} = V_{\text{piezo}} = Eh = g\varepsilon Yh$ .

We assume that the piezoelectric material is insulating and that the Schottky barrier between the piezoelectric material and the metal electrode can fully block the electron injection from the metal electrode back to the piezoelectric material. Thus, when the two electrodes are connected by an external load, the electrons will be driven to flow from one electrode to the other. These transferred electrons will generate another electric field, which will screen the original piezoelectric field generated by the strain. If the transferred charge amount from one electrode to the other electrode is  $Q$ , the voltage generated by those charges can be shown to be  $Q/C_0$ , where  $C_0$  is the capacitance between the two electrodes, which can be roughly estimated as a constant. Under short-circuit conditions, the total potential drop between the two electrodes is 0. Thus, we have  $V_{\text{piezo}} = Q_{\text{SC}}/C_0$ . The short-circuit transferred charge can be given by  $Q_{\text{SC}} = C_0 V_{\text{piezo}} = C_0 g\varepsilon Yh$ . Therefore the short-circuit current can be shown to be

$$I_{\text{SC}} = \frac{dQ_{\text{SC}}}{dt} = C_0 g Y h \frac{d\varepsilon}{dt}$$

In our experiment, the equivalent resistance of the piezoelectric nanogenerator (mainly contributed by the reverse-biased Schottky barrier) is  $\sim 80 \text{ G}\Omega$  when the bias is close to zero (at  $\sim 20 \text{ mV}$ ), which can be obtained from Fig. 3a. Because the largest load resistance used in power output measurement (Fig. 2) is  $2 \text{ G}\Omega$ , we can consider that the internal resistance of the piezoelectric nanogenerator to be infinite, and the Schottky barrier between the piezoelectric material and the metal electrode



can fully block electron injection from the metal electrode back to the piezoelectric material. Therefore the piezoelectric nanogenerator can be assumed to be purely capacitive. When the nanogenerator is connected with a load resistor, the equivalent circuit for the whole system is shown in Extended Data Fig. 4.

We first analyse a simple case when the mechanical motion is a pure harmonic vibration:  $\varepsilon = A \sin(\omega t)$ . Thus, from the above analysis,  $V_{\text{piezo}} = g\varepsilon Yh = gAYh \sin(\omega t)$ . And  $C_0$  can be assumed constant, as discussed above. The whole system is a linear time-invariant system and a phase method can be used to solve the system. The following equations for the output voltage and current with a load resistance of  $R$  can be obtained:

$$I_R = \frac{1}{(1/j\omega C_0) + R} gAYh = gAYh \frac{\omega C_0}{1 + \omega^2 R^2 C_0^2} (\omega RC_0 + j)$$

$$V_R = \frac{R}{(1/j\omega C_0) + R} gAYh = gAYh \frac{\omega RC_0}{1 + \omega^2 R^2 C_0^2} (\omega RC_0 + j)$$

Therefore

$$I_R(t) = gAYh \frac{\omega C_0}{1 + \omega^2 R^2 C_0^2} (\omega RC_0 \sin(\omega t) + \cos(\omega t))$$

$$V_R(t) = gAYh \frac{\omega RC_0}{1 + \omega^2 R^2 C_0^2} (\omega RC_0 \sin(\omega t) + \cos(\omega t))$$

The short-circuit current is then

$$I_{\text{SC}} = gAYh\omega C_0 \cos(\omega t)$$

Therefore the phase shift  $\theta$  between the short-circuit current and non-open-circuit voltage (finite external load) can be obtained as  $\theta = 90 - \arctan(1/\omega RC_0)$ . Therefore, when  $R = 0$  the phase shift between the output signal and the short-circuit current is 0, and when  $R = \infty$  the phase shift between the output signal and the short circuit current is  $90^\circ$ . For other cases, the phase difference is between  $0^\circ$  and  $90^\circ$ .

When the applied strain is not pure harmonic, because the applied strain is still a periodic signal it can be presented as a Fourier series:

$$\varepsilon = \sum_{k=1}^{\infty} A_1 \sin(k\omega t) + B_1 \cos(k\omega t)$$

Because the system is a still linear time-invariant system, the output is a summation of the output of each harmonic signal. From the similar analysis, we can still obtain a similar conclusion:

$$V_{\text{piezo}} = g\varepsilon Yh = gYh \left( \sum_{k=1}^{\infty} A_1 \sin(k\omega t) + B_1 \cos(k\omega t) \right)$$

$$I_R(t) = gYh \sum_{k=1}^{\infty} A_1 \left( \frac{k\omega C_0}{1 + k^2 \omega^2 R^2 C_0^2} (k\omega RC_0 \sin(k\omega t) + \cos(k\omega t)) \right)$$

$$+ B_1 \left( \frac{k\omega C_0}{1 + k^2 \omega^2 R^2 C_0^2} (k\omega RC_0 \cos(k\omega t) - \sin(k\omega t)) \right)$$

$$V_R(t) = gYh \sum_{k=1}^{\infty} A_1 \left( \frac{kR\omega C_0}{1 + k^2 \omega^2 R^2 C_0^2} (k\omega RC_0 \sin(k\omega t) + \cos(k\omega t)) \right)$$

$$+ B_1 \left( \frac{kR\omega C_0}{1 + k^2 \omega^2 R^2 C_0^2} (k\omega RC_0 \cos(k\omega t) - \sin(k\omega t)) \right)$$

When  $R = 0$ ,

$$I_{\text{SC}}(t) = gYh\omega C_0 \sum_{k=1}^{\infty} (A_1 k \cos(k\omega t) - B_1 k \sin(k\omega t))$$

When  $R = \infty$ ,

$$V_{\text{OC}}(t) = gYh \left( \sum_{k=1}^{\infty} A_1 \sin(k\omega t) + B_1 \cos(k\omega t) \right)$$

When  $R \ll 1/\omega C_0$ , it can be easily seen that  $I_R(t)$  converges to  $I_{\text{SC}}$ , and that when  $R \gg 1/\omega C_0$ ,  $V_R(t)$  converges to  $V_{\text{OC}}$ .

Consequently, on the basis of the above discussions for the two cases, there is a phase difference between the signals of the short-circuit current and non-open-circuit voltage. The difference depends on the internal impedance of the piezoelectric nanogenerator and the external load resistance. Similarly, there is also a difference

between the signals of the open-circuit voltage and the short-circuit current: the piezoelectric open-circuit voltage signal peaks when the strain reaches a maximum value, and the piezoelectric short-circuit current signal peaks when the strain rate reaches a maximum. It should be noted that the phase difference highlighted by black dashed lines was obtained by theoretical derivation for an ideal case; it was not measured experimentally because of the limitation in the experimental setup.

Moreover, the following relationship between acceleration  $a$ , maximum speed  $v_{\text{max}}$  and the total moving distance  $d$  (with maximum deflection  $D_{\text{max}}$ ) can be obtained:  $v_{\text{max}}^2 = 2aD_{\text{max}}/2 = aD_{\text{max}}$ , because  $\varepsilon \propto d$ , and

$$I_{\text{SC}} = \frac{dQ_{\text{SC}}}{dt} = C_0 gYh \frac{d\varepsilon}{dt}$$

$$I_{\text{SC,max}} = C_0 gYhkv_{\text{max}} = C_0 gYhkv_{\text{max}} \sqrt{aD_{\text{max}}}$$

Therefore, when we increase either the acceleration or the total moving distance  $D_{\text{max}}$ , the magnitude of the short-circuit current will increase. The hysteresis observed for the short-circuit current and the open-circuit voltage under strain can also be understood from the above discussion (Extended Data Fig. 5a).

**Estimation of power generation efficiency by MoS<sub>2</sub>.** The total mechanical deformation energy stored in the monolayer MoS<sub>2</sub> after being strained is calculated by  $W_{\text{M}} = LE\varepsilon^2/2$ , where  $E$  is the 2D elastic modulus of MoS<sub>2</sub> ( $130 \text{ N m}^{-1}$ )<sup>32</sup>,  $L$  is the distance between the two electrodes, and  $W$  is the width of the MoS<sub>2</sub> flake. The total electric energy generated through the piezoelectric polarization can be obtained as  $1.86 \times 10^{-14} \text{ J}$ . Therefore the corresponding energy conversion efficiency ( $\eta = W_{\text{E}}/W_{\text{M}}$ ) is found to be  $\sim 5.08\%$  for  $\varepsilon = 0.53\%$  ( $W_{\text{M}} = 3.66 \times 10^{-13} \text{ J}$ ).

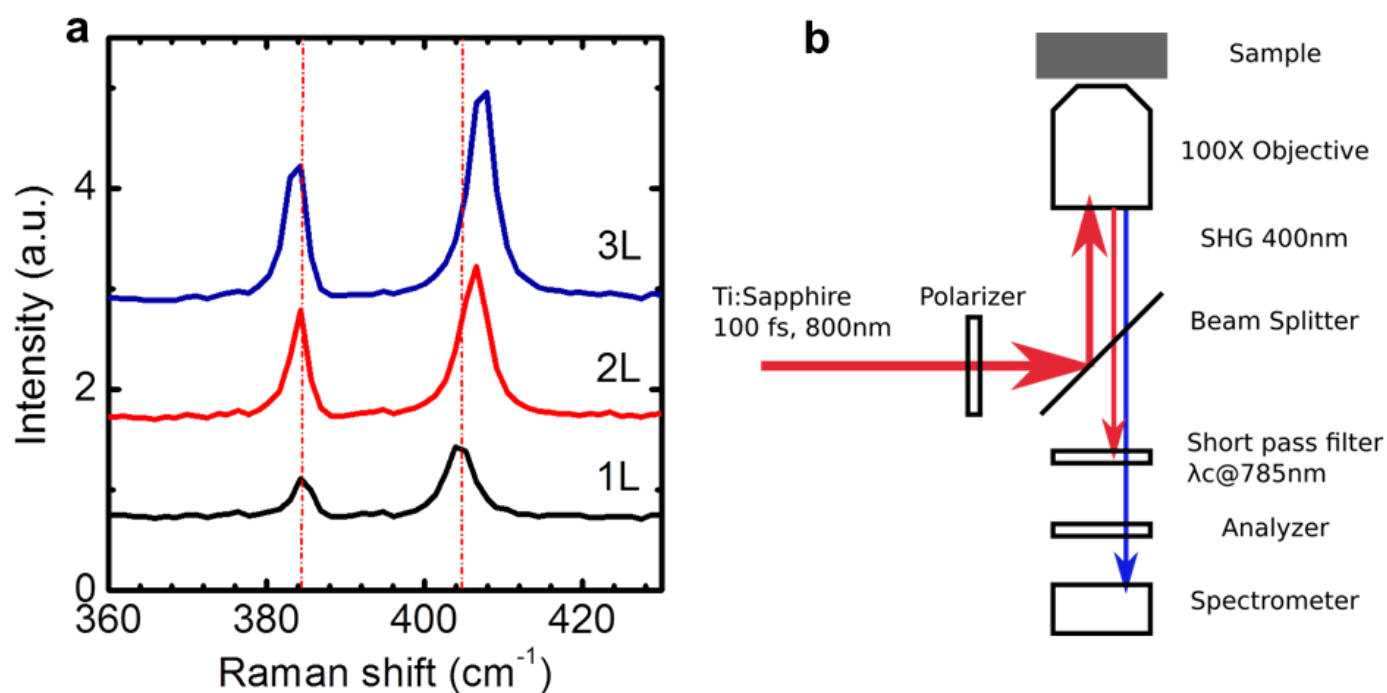
**Strain-modulated carrier transport in the MoS<sub>2</sub> piezotronic effect and piezoresistive effect.** The observed difference in conduction behaviour under strain between monolayer and bilayer/bulk MoS<sub>2</sub> is attributed to their fundamental distinction in crystal symmetries. It has been theoretically predicted that MoS<sub>2</sub> with odd number of layers (point group  $D_{3h}$ ) will exhibit intrinsic piezoelectric property as a result of its lack of centrosymmetry in crystal structure, whereas piezoelectricity will vanish in centrosymmetrically structured MoS<sub>2</sub> with an even number of layers (point group  $D_{6h}$ )<sup>9,10,16</sup>. Resulting from changes in band structure, charge carrier density or the density of states in the conduction band of strained semiconductor materials, the piezoresistive effect is symmetrical on the two end contacts and has no polarity; this will not produce the function of a transistor. Piezoresistance is a common feature of semiconductors such as Si and GaAs and is not limited to piezoelectric semiconductors. The strain-induced modification of band structures in MoS<sub>2</sub> has recently been reported with a small range of compressive strain ( $<2\%$ ) increasing and tensile strain decreasing the bandgap for MoS<sub>2</sub>, as a result of the change in orbital overlap and hybridization by strain<sup>33–36</sup>. The trigonal-prismatic coordination between the molybdenum and sulphur atoms and the absence of inversion symmetry give rise to piezoelectric polarization in strained monolayer MoS<sub>2</sub>, with the polarization charges induced at the zigzag edges. In general, the negative piezoelectric polarization charges, and hence the negative piezopotential induced at the semiconductor side near the interface of the local contact formed between the metal electrode and an  $n$ -type semiconductor, can repel the electrons from the interface, resulting in a further depleted interface and increased local barrier heights, whereas the positive piezoelectric polarization charges and hence the positive piezopotential created at the semiconductor side near the interface can attract the electrons towards the interface, resulting in a less depleted interface and hence decreased local barrier heights. The strain-induced polarization charges can therefore directly affect the local contacts at the metal–MoS<sub>2</sub> interfaces by exerting substantial influences on the concentration and distribution of free carriers in MoS<sub>2</sub> as well as on the modulation of electronic charge in interface states or metal. The observed anisotropic changes in current transport with applied strain are associated with variations in Schottky barrier height (SBH) tuned by both the strain-induced change in band structure and piezoelectric polarization charges at a reverse-biased Pd–MoS<sub>2</sub> barrier<sup>37</sup>. The contributions to changes in SBH at source and drain contacts from the band structure effect share the same polarity: uniaxial compressive ( $<2\%$ ) or tensile strain in the direction along armchair edge increases or decreases the bandgap for monolayer MoS<sub>2</sub> (refs 33–36), whereas the modulation of SBHs at both contacts due to the polarization-induced surface charges and the corresponding adjustment of electronic states at the interface possess the opposite polarities owing to the polarity of induced piezoelectric charges<sup>37</sup>. These piezoelectric polarization charges can effectively modulate the local contact characteristics through an internal field, depending on doping type, carrier density and the crystallographic orientation of the piezoelectric semiconductor material as well as on the polarity of the applied strain. Consequently, the transport of charge carriers across the metal–semiconductor contact can be effectively modulated by the piezoelectric polarization charges, which can be controlled by varying the magnitude and polarity of the externally applied strain. The modulation or gating of the charge transport across the interface by the strain-induced polarization charges is the core of piezotronics. The gauge factor

$([\Delta I(\varepsilon)/I(0)]/\Delta\varepsilon)$  of the monolayer MoS<sub>2</sub> device for strain sensing has also been characterized; the highest value is  $\sim 760$ , which exceeds the values of conventional metal strain gauges ( $\sim 1-5$ ), a state-of-the-art silicon strain sensor ( $\sim 200$ ) and a graphene strain sensor ( $\sim 300$ )<sup>26</sup>, suggesting the potential for using monolayer MoS<sub>2</sub> in highly sensitive strain sensing. The higher gauge factor in a monolayer MoS<sub>2</sub> strain sensor than those for graphene, bilayer MoS<sub>2</sub> ( $\sim 230$ ) and bulk MoS<sub>2</sub> ( $\sim 200$ ) is attributed to enhancement by piezoelectric polarization. However, we are not sure why bilayer devices show slightly higher gauge factors than bulk samples. We think that relative change in carrier density and/or mobility by strain is more significant in bilayer devices than in bulk samples, which may be related to the difference between the band structures of bilayer and bulk MoS<sub>2</sub>. Last, the effect from substrate contact with the bilayer structure may also contribute to the observed large piezoresistance of the device. Moreover, the much smaller thickness and original conductivity in bilayer devices may also have a role here. Both theoretical calculations (for example, tight binding) and experimental characterizations are required for a better understanding of the piezoresistive effect in bilayer MoS<sub>2</sub>.

**Internal screening of piezoelectric polarization charges in MoS<sub>2</sub> by free carriers.** The partial screening of strain-induced polarization charges by free carriers has been widely observed for conventional piezoelectric semiconductors such as ZnO (refs 38, 39). Because MoS<sub>2</sub> has a relatively small bandgap ( $\sim 1.8$  eV), it is anticipated that the internal screening of piezoelectric charges should exist. Considering the *n*-type doping characteristics of MoS<sub>2</sub>, the positive polarization charges induced by strain will be partly screened by the free electrons in MoS<sub>2</sub>, whereas the negative polarization charges will be preserved. Therefore it is still possible to observe piezoelectric power output and piezotronic transport. The trend in band diagrams shown in Fig. 3c is still valid if the MoS<sub>2</sub> is not heavily doped, except that

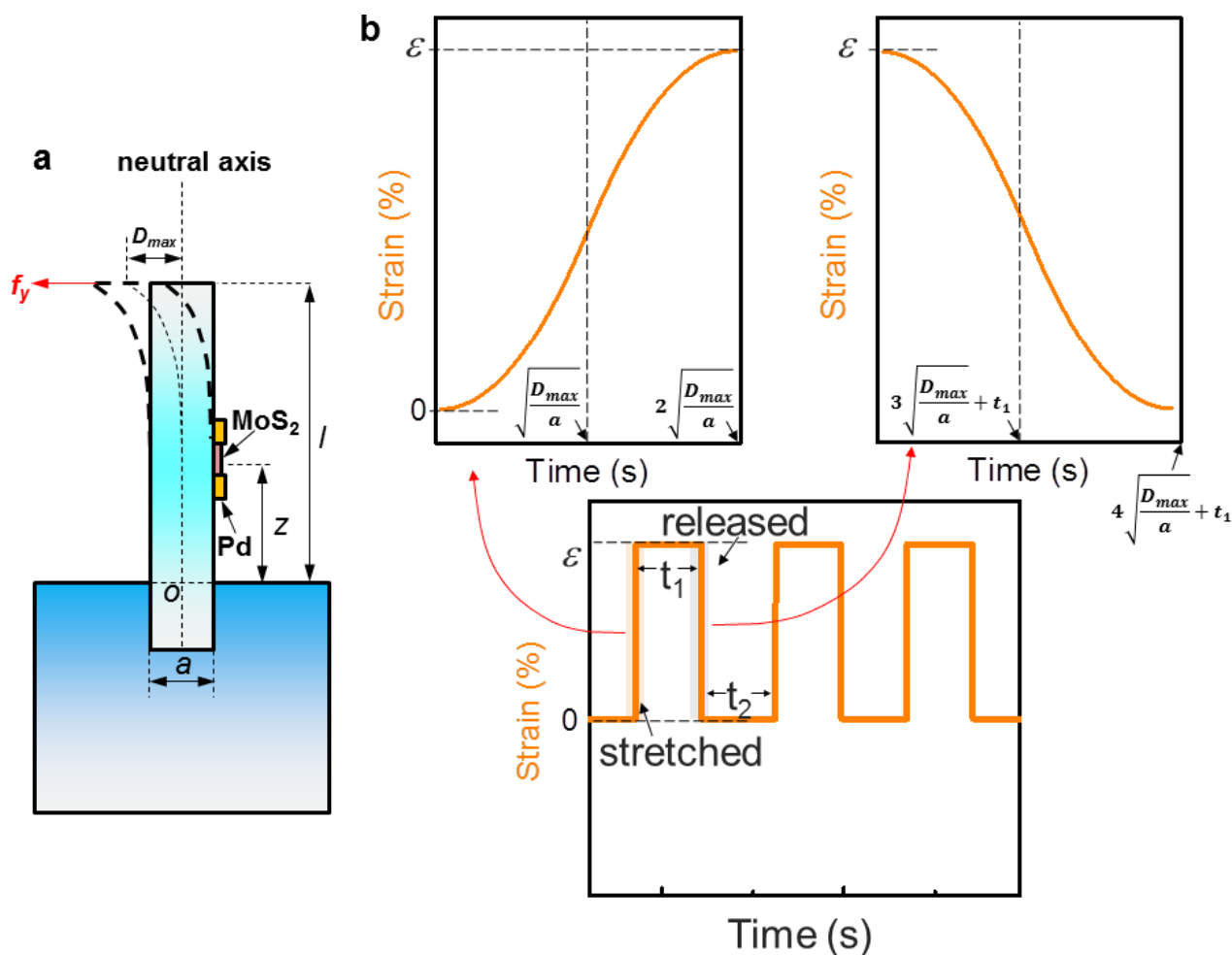
the changes in SBH induced by positive piezoelectric charges will be decreased. Considering the much smaller dimensions of 2D materials than those of conventional piezoelectric semiconductors such as ZnO nanowires, the carrier density or even conductivity type in monolayer MoS<sub>2</sub> may be affected or modulated in a more efficient or sensitive way by substitutional doping at both the Mo and the S sites and by the adsorption of charged molecules. More in-depth investigations are therefore needed in both theory and experiments to quantify the effect of internal screening of piezoelectric polarization in 2D piezoelectric semiconductors.

31. Landau, L. D., Lifshits, E. M., Kosevich, A., d. M. & Pitaevskii, L. P. *Theory of Elasticity* 3rd English edn (Pergamon, 1986).
32. Cooper, R. C. *et al.* Nonlinear elastic behavior of two-dimensional molybdenum disulfide. *Phys. Rev. B* **87**, 035423 (2013).
33. He, K., Poole, C., Mak, K. F. & Shan, J. Experimental demonstration of continuous electronic structure tuning via strain in atomically thin MoS<sub>2</sub>. *Nano Lett.* **13**, 2931–2936 (2013).
34. Conley, H. *et al.* Bandgap engineering of strained monolayer and bilayer MoS<sub>2</sub>. *Nano Lett.* **13**, 3626–3630 (2013).
35. Hui, Y. Y. *et al.* Exceptional tunability of band energy in a compressively strained trilayer MoS<sub>2</sub> sheet. *ACS Nano* **7**, 7126–7131 (2013).
36. Johari, P. & Shenoy, V. B. Tuning the electronic properties of semiconducting transition metal dichalcogenides by applying mechanical strains. *ACS Nano* **6**, 5449–5456 (2012).
37. Zhang, Y., Liu, Y. & Wang, Z. L. Fundamental theory of piezotronics. *Adv. Mater.* **23**, 3004–3013 (2011).
38. Gao, Y. F. & Wang, Z. L. Equilibrium potential of free charge carriers in a bent piezoelectric semiconductive nanowire. *Nano Lett.* **9**, 1103–1110 (2009).
39. Hu, Y. F., Lin, L., Zhang, Y. & Wang, Z. L. Replacing a battery by a nanogenerator with 20 V output. *Adv. Mater.* **24**, 110–114 (2012).



**Extended Data Figure 1 | Raman spectrum of MoS<sub>2</sub> flakes and setup for SHG measurement.** **a**, Raman spectrum of MoS<sub>2</sub> flakes with different layer numbers. **b**, Experimental setup for the SHG measurement.





**c** Stretching

Name	Value
Position	10 mm
Acceleration	5 m/s <sup>2</sup>
Deceleration	5 m/s <sup>2</sup>

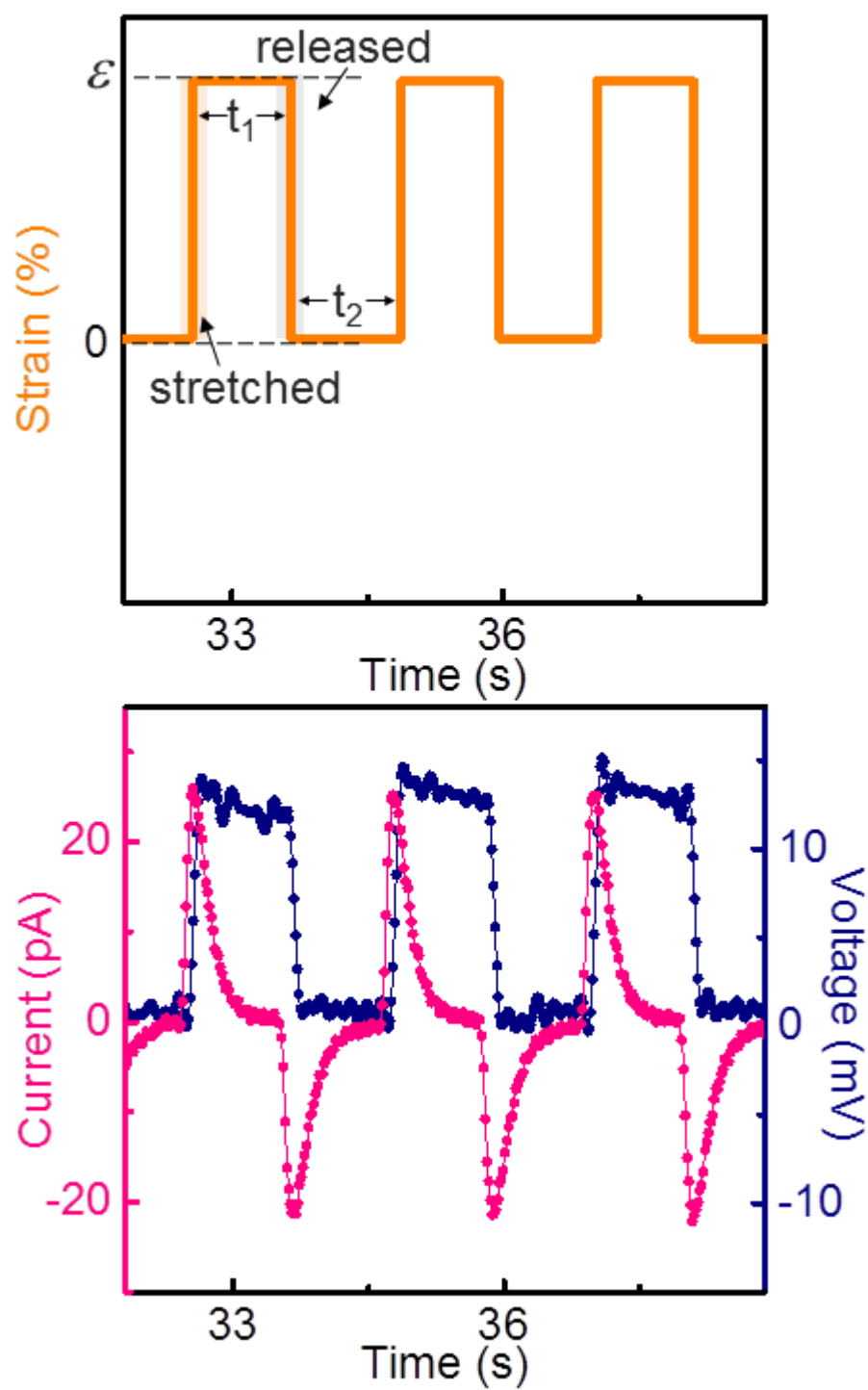
Releasing

Name	Value
Position	0 mm
Acceleration	5 m/s <sup>2</sup>
Deceleration	5 m/s <sup>2</sup>

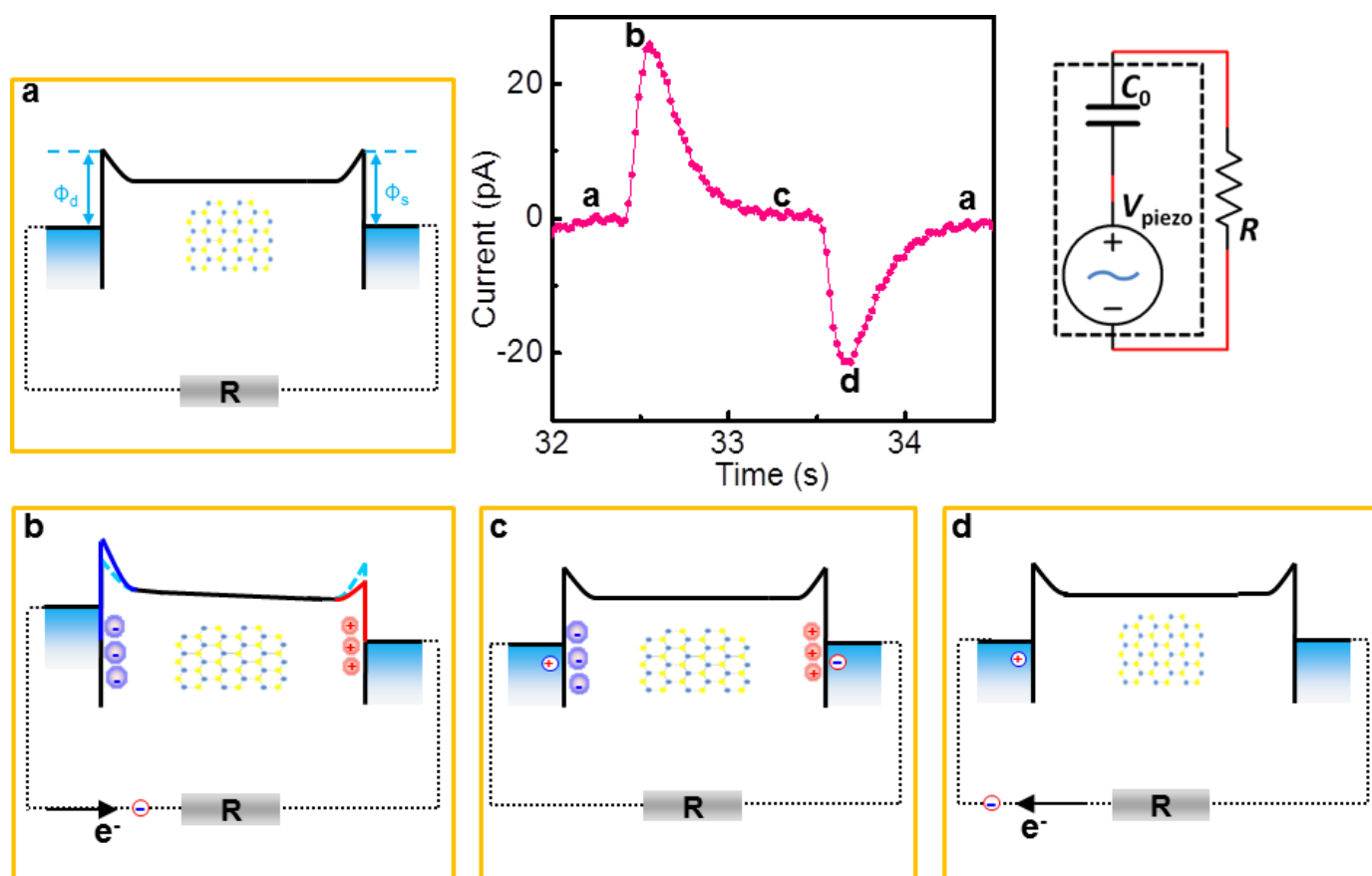
Holding

Name	Value
Rise Position Wait Time	1000 ms
Fall Position Wait Time	1000 ms

**Extended Data Figure 2 | Mechanical strain applied to MoS<sub>2</sub> device.** **a**, Schematic drawing for estimating strain in MoS<sub>2</sub> device. **b**, Schematic plot of strain driving signal from linear motor. **c**, Typical configurations for linear motor.



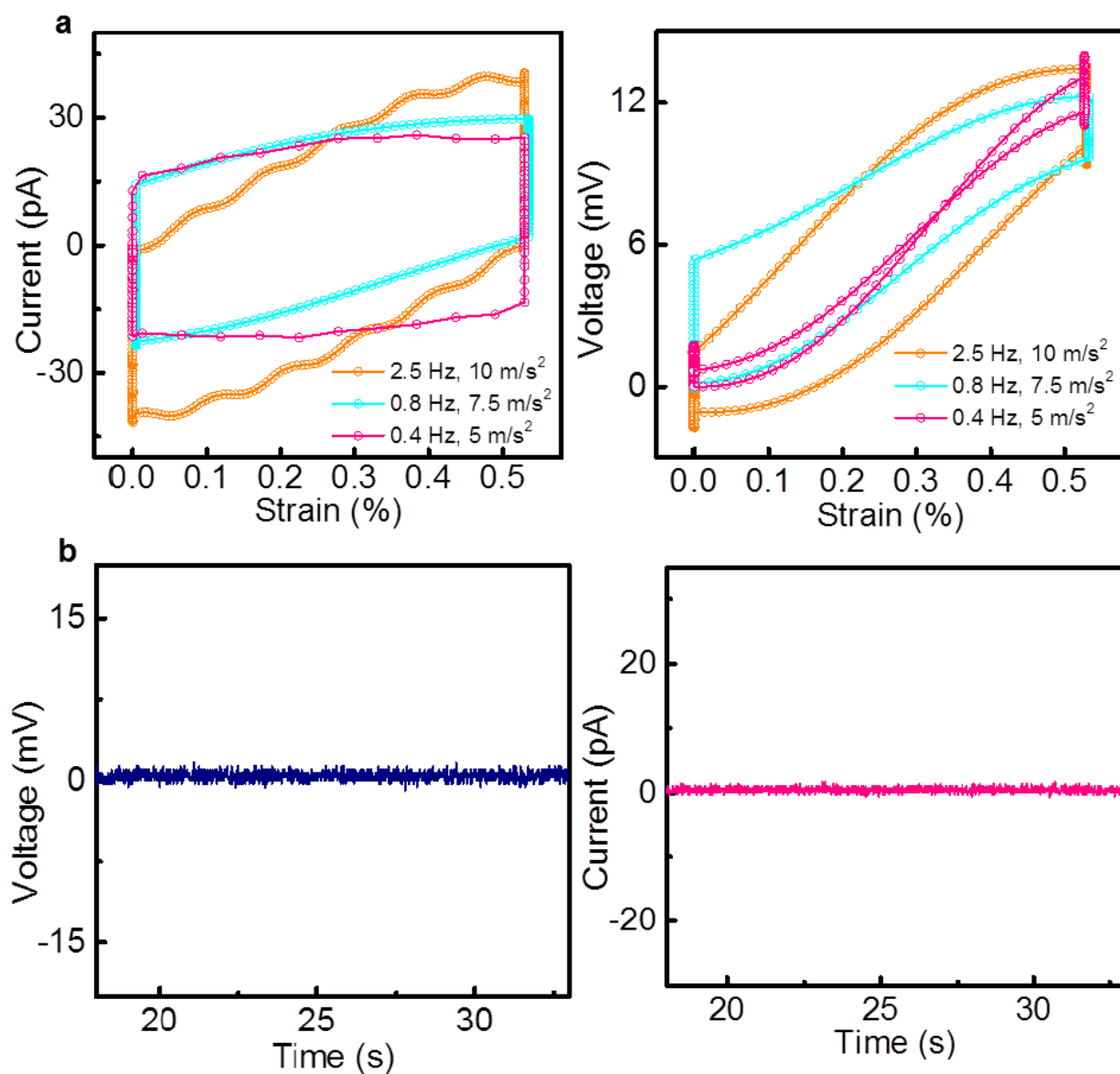
Extended Data Figure 3 | Piezoelectric open-circuit voltage and short-circuit current.



Extended Data Figure 4 | Mechanism of electrical power generation in single-layer MoS<sub>2</sub> due to the flow of electrons in external load driven by

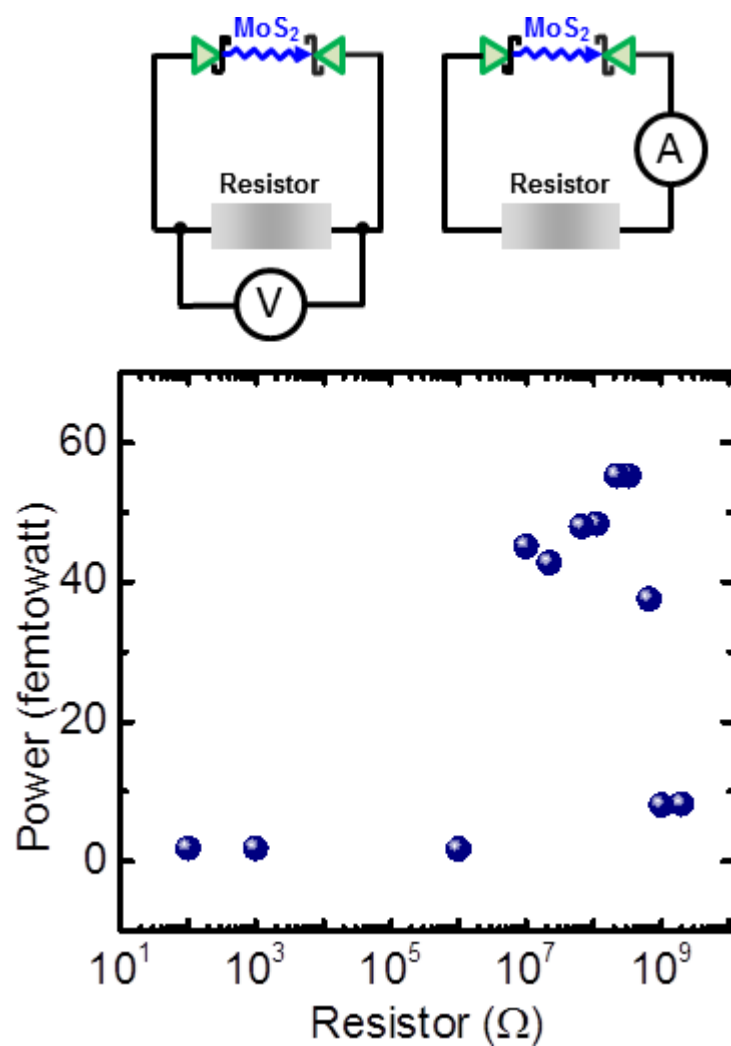
piezoelectric polarization charges. The equivalent circuit of the piezoelectric nanogenerator is also shown.



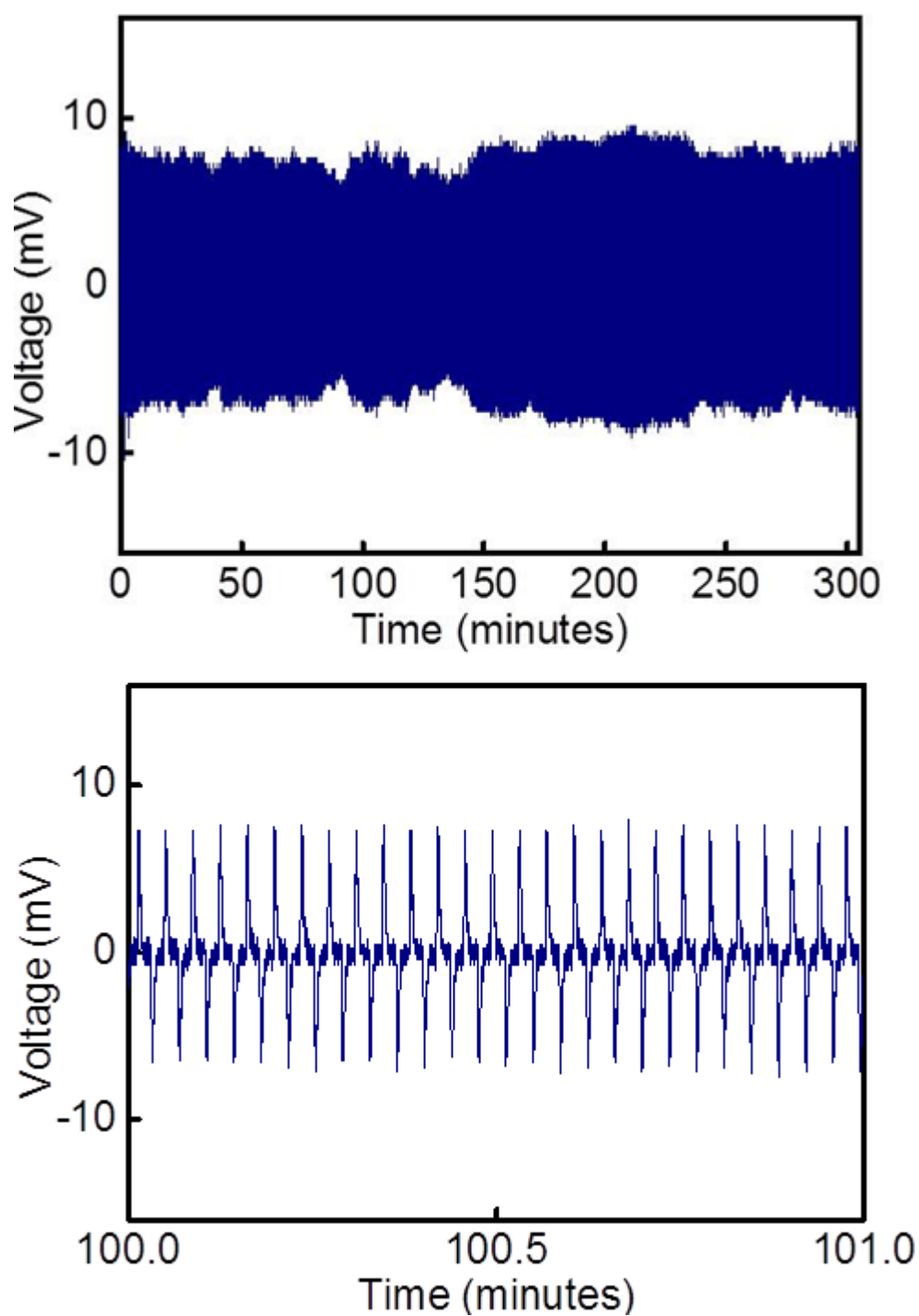


**Extended Data Figure 5 | Piezoelectric output of MoS<sub>2</sub> device with different strain parameters.** **a**, Short-circuit current–strain and open-circuit voltage–strain hysteresis loops. Hold time  $t_1 = t_2 = 1$  s and acceleration  $a = 5 \text{ m s}^{-2}$  for the curve of 0.4 Hz; hold time  $t_1 = t_2 = 0.5$  s and acceleration

$a = 7.5 \text{ m s}^{-2}$  for 0.8 Hz; hold time  $t_1 = t_2 = 0.1$  s and acceleration  $a = 10 \text{ m s}^{-2}$  for 2.5 Hz. **b**, Electrical outputs from bare PET substrate without single-layer MoS<sub>2</sub> under periodic strain (0.53%).



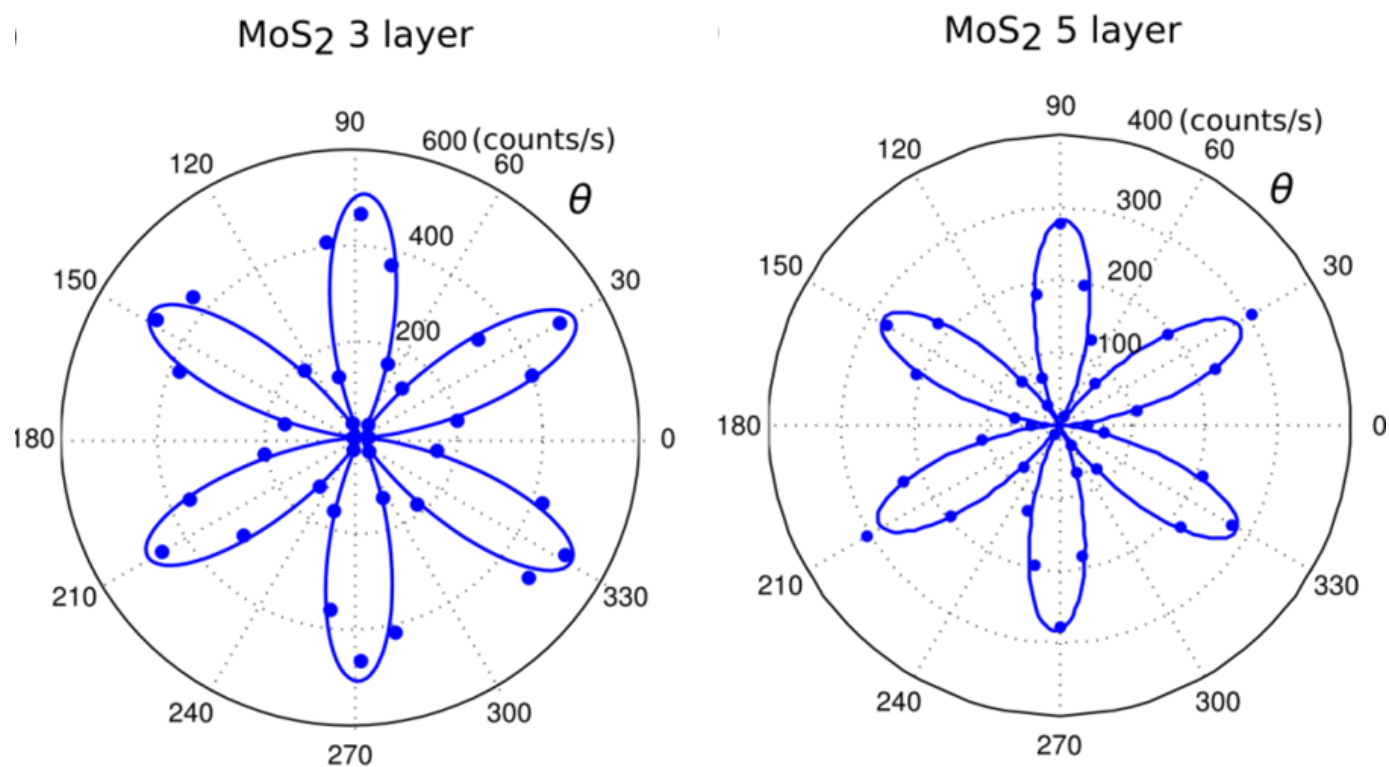
Extended Data Figure 6 | Circuit connection for measuring the power outputs on the external load and power delivered to the load at 0.53% strain.



**Extended Data Figure 7 | Stability test of voltage output from single-layer MoS<sub>2</sub> device.** The frequency of 0.5 Hz was held for 300 min. The results

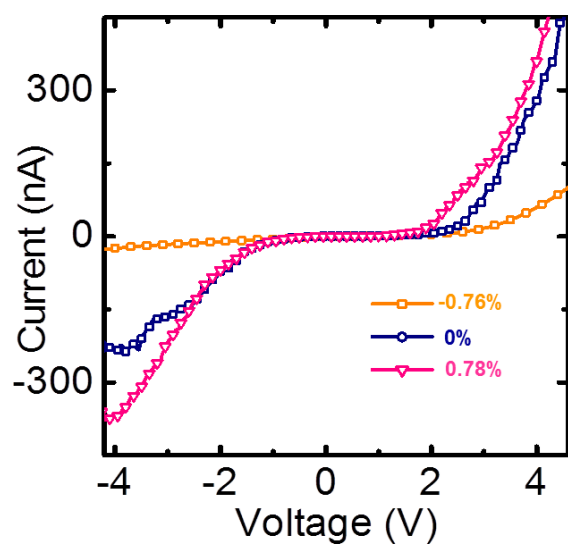
demonstrate good stability of the device in mechanical energy harvesting for prolonged periods.



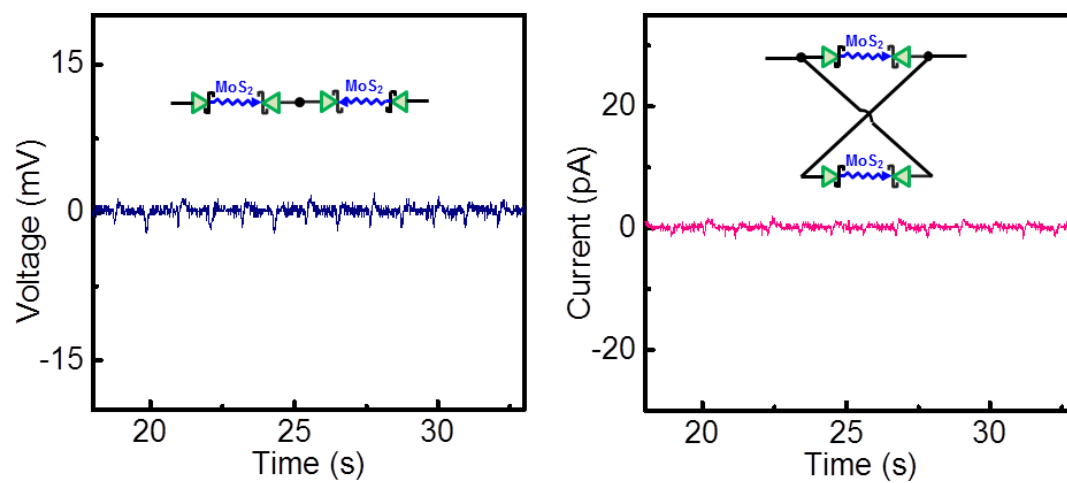


**Extended Data Figure 8 | Angular dependence of SHG intensity (perpendicular component) for three-layer and five-layer MoS<sub>2</sub>.** Samples of even layers (two, four and six layers) give vanishing SHG intensity regardless of

their crystallographic orientation.  $\theta$  denotes the angle between the fundamental light polarization and the mirror plane of the lattice.



Extended Data Figure 9 | Transport characteristics of bulk device under different uniaxial strains.



Extended Data Figure 10 | Electrical outputs when CVD devices 1 and 2 are destructively connected.



# Identification of an iridium-containing compound with a formal oxidation state of IX

Guanjun Wang<sup>1</sup>, Mingfei Zhou<sup>1</sup>, James T. Goettel<sup>2</sup>, Gary J. Schrobilgen<sup>2</sup>, Jing Su<sup>3</sup>, Jun Li<sup>3</sup>, Tobias Schlöder<sup>4</sup> & Sebastian Riedel<sup>4,5</sup>

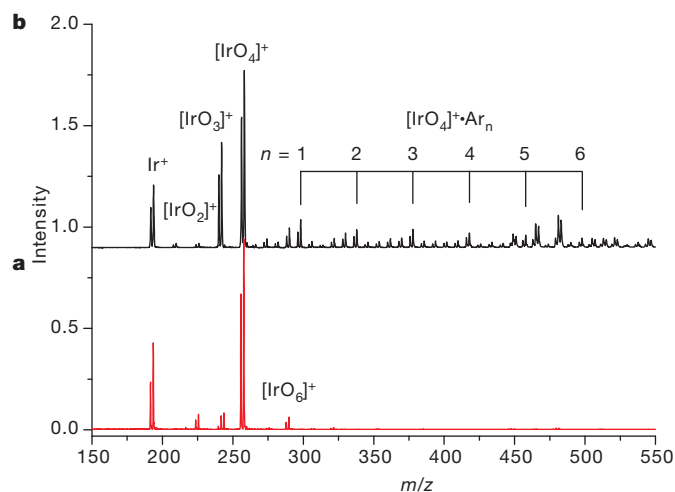
One of the most important classifications in chemistry and within the periodic table is the concept of formal oxidation states<sup>1–4</sup>. The preparation and characterization of compounds containing elements with unusual oxidation states is of great interest to chemists<sup>5</sup>. The highest experimentally known formal oxidation state of any chemical element is at present VIII<sup>2–4</sup>, although higher oxidation states have been postulated<sup>6,7</sup>. Compounds with oxidation state VIII include several xenon compounds<sup>8</sup> (for example XeO<sub>4</sub> and XeO<sub>3</sub>F<sub>2</sub>) and the well-characterized species RuO<sub>4</sub> and OsO<sub>4</sub> (refs 2–4). Iridium, which has nine valence electrons, is predicted to have the greatest chance of being oxidized beyond the VIII oxidation state<sup>1</sup>. In recent matrix-isolation experiments, the IrO<sub>4</sub> molecule was characterized as an isolated molecule in rare-gas matrices<sup>9</sup>. The valence electron configuration of iridium in IrO<sub>4</sub> is 5d<sup>1</sup>, with a formal oxidation state of VIII. Removal of the remaining *d* electron from IrO<sub>4</sub> would lead to the iridium tetroxide cation ([IrO<sub>4</sub>]<sup>+</sup>), which was recently predicted to be stable<sup>10</sup> and in which iridium is in a formal oxidation state of IX. There has been some speculation about the formation of [IrO<sub>4</sub>]<sup>+</sup> species<sup>11,12</sup>, but these experimental observations have not been structurally confirmed. Here we report the formation of [IrO<sub>4</sub>]<sup>+</sup> and its identification by infrared photodissociation spectroscopy. Quantum-chemical calculations were carried out at the highest level of theory that is available today, and predict that the iridium tetroxide cation, with a T<sub>d</sub>-symmetrical structure and a d<sup>0</sup> electron configuration, is the most stable of all possible [IrO<sub>4</sub>]<sup>+</sup> isomers.

Iridium oxide cations were generated in the gas phase using a pulsed-laser vaporization/supersonic-expansion source and were studied by infrared photodissociation spectroscopy in the 850–1,600 cm<sup>−1</sup> region as described previously<sup>13</sup>. A typical mass spectrum of the iridium oxide cations produced with O<sub>2</sub>-seeded helium is shown in Fig. 1a. The spectrum consists of a progression of peaks having different masses that correspond to different [IrO<sub>*x*</sub>]<sup>+</sup> species with up to six oxygen atoms. Enhanced abundance in the mass spectrum was found for [IrO<sub>4</sub>]<sup>+</sup>, and the preferential formation of this cation indicates increased stability. Because the dissociation energies of the [IrO<sub>4</sub>]<sup>+</sup> cations are significantly greater than the infrared photon energies in the Ir=O and O–O stretching frequency region (the infrared photons in the 900–1,200 cm<sup>−1</sup> region have energies in the range of 10.8–14.4 kJ mol<sup>−1</sup>; Supplementary Information), the method of rare-gas atom predissociation is employed to obtain the infrared spectra for these molecules<sup>14–17</sup>. When argon was used instead of helium, [IrO<sub>4</sub>]<sup>+</sup>·Ar<sub>*n*</sub> ions (*n* = 1, 2 and larger; the dot denotes a weak bond) were produced (Fig. 1b) and the mass peaks corresponding to the [<sup>193</sup>IrO<sub>4</sub>]<sup>+</sup>·Ar<sub>*n*</sub> isotopomers were selected for photodissociation. When the infrared laser is on resonance with one of the vibrational fundamentals of an [IrO<sub>4</sub>]<sup>+</sup>·Ar<sub>*n*</sub> complex, the latter photodissociates by eliminating an argon atom. The resulting predissociation infrared spectra of [IrO<sub>4</sub>]<sup>+</sup> are shown in Fig. 2.

The experimental spectrum of [IrO<sub>4</sub>]<sup>+</sup>·Ar (Fig. 2a) consists of five absorptions at 936, 944, 966, 1,047 and 1,054 cm<sup>−1</sup>, respectively, (Table 1),

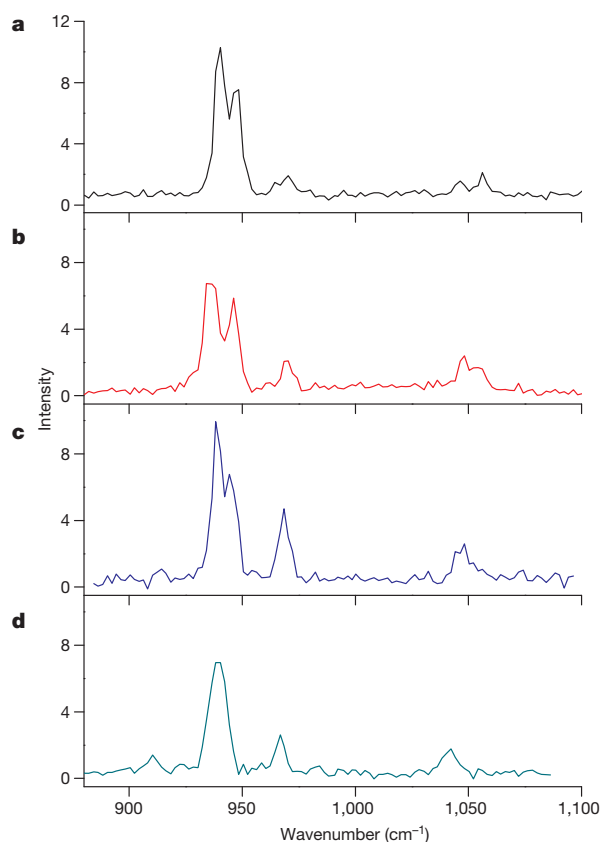
indicating that more than one isomer is experimentally observed, because any [IrO<sub>4</sub>]<sup>+</sup> structure should at most have four vibrational fundamentals in the Ir=O and O–O stretching regions. Experiments with different time delays between expansion from the pulsed valve and vaporization (Methods), and with different stagnation pressures, suggest that the bands at 936 and 944 cm<sup>−1</sup> are due to the same species, whereas the other three absorptions are caused by other isomers. All bands were shifted to lower wavenumbers in the experiments using <sup>18</sup>O<sub>2</sub> (Supplementary Information), and the observed frequency shifts with <sup>16</sup>O/<sup>18</sup>O isotopic ratios in the range of 1.050–1.056 suggest that these bands originate from Ir=O or O–O stretching vibrations.

Recent quantum-chemical calculations show that three low-energy isomers are possible for a cation of [IrO<sub>4</sub>]<sup>+</sup> stoichiometry<sup>10</sup>. These cations have been reinvestigated by more accurate *ab initio* coupled-cluster calculations with single and double excitations and perturbative-triples corrections (CCSD(T)), as well as by *ab initio* multi-reference-based complete active space perturbation theory (CASPT2) calculations (Methods). On the basis of these calculations, the wavenumber of the O–O stretching mode of the superoxide complex is predicted to be 1,486.3 cm<sup>−1</sup> at density functional level with inclusion of dispersion corrections (B3LYP-D3) or 1,458.7 cm<sup>−1</sup> (CCSD(T)) with appreciable infrared intensity. Hence, this superoxide complex [(η<sup>1</sup>-O<sub>2</sub>)Ir<sup>VI</sup>O<sub>2</sub>]<sup>+</sup> (C<sub>s</sub> symmetry, <sup>3</sup>A'' ground state) can be ruled out because of the absence of any observed band in the experimental spectra above 1,100 cm<sup>−1</sup>. The observed absorptions probably come from the side-on O<sub>2</sub> complex [(η<sup>2</sup>-O<sub>2</sub>)Ir<sup>VII</sup>O<sub>2</sub>]<sup>+</sup> (C<sub>2v</sub>, <sup>1</sup>A<sub>1</sub>) and the tetroxide complex [Ir<sup>IX</sup>O<sub>4</sub>]<sup>+</sup> (T<sub>d</sub>, <sup>1</sup>A<sub>1</sub>), where the



**Figure 1 | Mass spectra of the iridium oxide cations.** The cations are produced by pulsed-laser vaporization of an iridium metal target in an expansion of helium (a) or argon (b) seeded by dioxygen. The isotopic splitting of iridium can clearly be resolved with the relative peak areas matching the natural abundance isotopic distribution (<sup>191</sup>Ir, 37.3%; <sup>193</sup>Ir, 62.7%). *m/z*, mass/charge ratio; intensity is shown in arbitrary units.

<sup>1</sup>Collaborative Innovation Center of Chemistry for Energy Materials, Department of Chemistry, Shanghai Key Laboratory of Molecular Catalysts and Innovative Materials, Fudan University, Shanghai 200433, China. <sup>2</sup>Department of Chemistry, McMaster University, 1280 Main Street West, Hamilton, Ontario L8S 4M1, Canada. <sup>3</sup>Department of Chemistry and Key Laboratory of Organic Optoelectronics and Molecular Engineering of Ministry of Education, Tsinghua University, Beijing 100084, China. <sup>4</sup>Institut für Anorganische und Analytische Chemie, Albert-Ludwigs Universität Freiburg, Albertstrasse 21, D-79104 Freiburg im Breisgau, Germany. <sup>5</sup>Institut für Chemie und Biochemie – Anorganische Chemie, Freie Universität Berlin, Fabeckstrasse 34–36, D-14195 Berlin, Germany.



**Figure 2 | Infrared photodissociation spectra of the  $[\text{IrO}_4]^+ \cdot \text{Ar}_n$  ( $n = 1-4$ ) cations.** The spectra are measured by monitoring the Ar photodissociation channel: **a**,  $[\text{IrO}_4]^+ \cdot \text{Ar}$ ; **b**,  $[\text{IrO}_4]^+ \cdot \text{Ar}_2$ ; **c**,  $[\text{IrO}_4]^+ \cdot \text{Ar}_3$ ; **d**,  $[\text{IrO}_4]^+ \cdot \text{Ar}_4$ . Intensity is shown as the yield of fragmentation ions normalized to the parent ion signal in percentage.

latter is by far the most stable isomer according to high-level (coupled-cluster) calculations (Fig. 3).

Following the general rules for the determination of formal oxidation states, the experimentally identified cation  $[\text{IrO}_4]^+$  can be viewed as an Ir(IX) species. This assignment is in line with the well-known isoelectronic  $\text{OsO}_4$  molecule, in which osmium undoubtedly exists in its oxidation state VIII. A more detailed discussion of the assignment of oxidation states based on natural population analysis and molecular orbitals is provided in Methods.

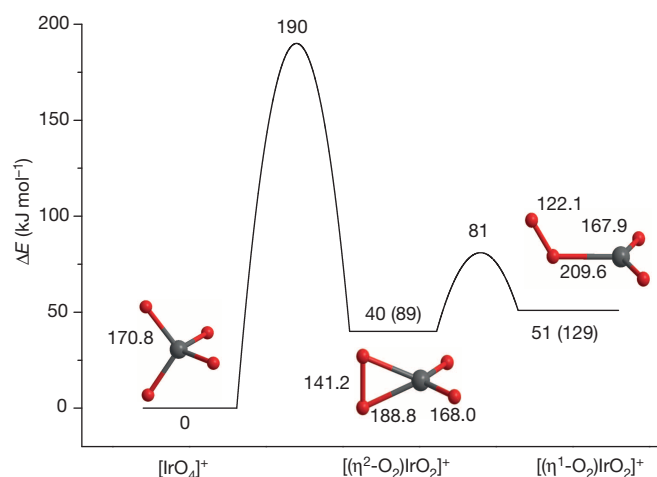
**Table 1 | Observed and calculated vibrational frequencies of the  $[\text{IrO}_4]^+$  and the face-coordinated  $\text{C}_{3v}$   $[\text{IrO}_4]^+ \cdot \text{Ar}$  isomers**

Isomer	$[\text{IrO}_4]^+$ ( $\text{cm}^{-1}$ )		$[\text{IrO}_4]^+ \cdot \text{Ar}$ ( $\text{cm}^{-1}$ )		Exptl	Mode*
	CCSD(T)†	B3LYP†	CCSD(T)†	B3LYP†		
$[\text{IrO}_4]^+$	916.4	1,009.8	916.7	1,008.9	936	Sym. str. ( $A_1$ )
	(0)	(0)	(0)	(0)		
	963.0	1,008.5	963.3	1,007.5		Antisym. str. ( $T_2$ )
	(35)	(39)	(13)	(40)		
	963.0	1,008.5	963.5	1,009.0	944	Antisym. str. ( $T_2$ )
$[(\eta^2\text{-O}_2)\text{IrO}_2]^+$	(35)	(39)	(11)	(39)		
	963.0	1,008.5	963.5	1,009.0		Antisym. str. ( $T_2$ )
	(35)	(39)	(11)	(39)		
	954.9	1,026.8	—	1,025.0	966	OlrO antisym. str. ( $B_1$ )
	(55)	(70)	—	(72)		
$[(\eta^1\text{-O}_2)\text{IrO}_2]^+$	1,000.6	1,062.7	—	1,061.9	1,047	OlrO sym. str. ( $A_1$ )
	(11)	(12)	—	(19)		
	996.5	1,032.8	—	1,027.9	1,054	O—O str. ( $A_1$ )
	(56)	(60)	—	(59)		

Infrared intensities are listed in parentheses in  $\text{km mol}^{-1}$ .

\* Mode descriptions and symmetry labels for the free cations.

† Calculated at CCSD(T)/aug-cc-pVTZ-PP and B3LYP-D3/aug-cc-pVTZ-PP levels. The frequencies of the edge-coordinated ( $\eta^2$ )  $\text{C}_{2v}$  structure are predicted at 916.8  $\text{cm}^{-1}$  (0  $\text{km mol}^{-1}$ ,  $A_1$ ), 963.0  $\text{cm}^{-1}$  (12  $\text{km mol}^{-1}$ ,  $A_1$ ), 963.4  $\text{cm}^{-1}$  (10  $\text{km mol}^{-1}$ ,  $B_2$ ), 963.6  $\text{cm}^{-1}$  (12  $\text{km mol}^{-1}$ ,  $B_1$ ) at the CCSD(T) level. All experiments have been replicated more than ten times.



**Figure 3 | Optimized structures and energetic ordering of the different  $[\text{IrO}_4]^+$  isomers.** The structures are optimized at the CCSD(T) level. The barriers of the interconversion reactions are also shown. The relative energies are calculated at the SO-DFT/B3LYP and CCSD(T) (in parentheses) levels, with the energy of the most stable isomer ( $[\text{IrO}_4]^+$ ) set to zero. The symmetries and ground states are as follows:  $[\text{IrO}_4]^+$  ( $T_d$ ,  $^1A_1$ ),  $[(\eta^2\text{-O}_2)\text{IrO}_2]^+$  ( $C_{2v}$ ,  $^1A_1$ ),  $[(\eta^1\text{-O}_2)\text{IrO}_2]^+$  ( $C_s$ ,  $^3A''$ ). Bond lengths are in pm; energies are in  $\text{kJ mol}^{-1}$ . Red, oxygen atoms; grey, iridium atoms.

The second most stable isomer of  $[\text{IrO}_4]^+$  stoichiometry is the  $\text{C}_{2v}$ -symmetrical side-on complex  $[(\eta^2\text{-O}_2)\text{Ir}^{\text{VII}}\text{O}_2]^+$ , which was also calculated to have a closed-shell electron configuration. All three isomers were calculated to be true minima on the potential energy surface and are energetically separated by 89 and 40  $\text{kJ mol}^{-1}$ , respectively, at the CCSD(T) level (Fig. 3). Our calculations show that their energetic ordering is not affected by weak argon coordination (Supplementary Information). The barrier for the  $[(\eta^1\text{-O}_2)\text{IrO}_2]^+ \rightarrow [(\eta^2\text{-O}_2)\text{IrO}_2]^+$  conversion was calculated to be only about 30  $\text{kJ mol}^{-1}$  using density functional theory with relativistic spin-orbit coupling effects (SO-DFT/B3LYP; Fig. 3 and Supplementary Information). The barrier for the  $[(\eta^2\text{-O}_2)\text{IrO}_2]^+ \rightarrow [\text{IrO}_4]^+$  reaction lies at 150  $\text{kJ mol}^{-1}$  at the SO-DFT/B3LYP level (Fig. 3 and Supplementary Information). This barrier is remarkably high, but nevertheless is plausible because the isomerization reaction involves the electron transfer and cleavage of the relatively strong peroxide O—O bond<sup>18</sup>. The  $[\text{Ir}(\eta^2\text{-O}_2)]^+ \rightarrow [\text{IrO}_2]^+$  isomerization reaction was computed to have a comparable barrier<sup>19</sup>.

From comprehensive analysis and quantum-chemical calculations, we assign the 936 and 944  $\text{cm}^{-1}$  bands to the antisymmetric iridium oxide stretching modes of the argon-tagged iridium tetroxide cation (Table 1 and Fig. 2a). Only the triply degenerate antisymmetric stretching mode ( $T_2$  mode) is infrared-active in the tetrahedral  $[\text{IrO}_4]^+$  cation. In the experimental spectrum of  $[\text{IrO}_4]^+ \cdot \text{Ar}$ , we identified two bands separated by 8  $\text{cm}^{-1}$  using argon-atom tagging. Our theoretical calculations on all three possible argon coordination modes to  $[\text{IrO}_4]^+$ , namely face coordination ( $\eta^3$ ,  $\text{C}_{3v}$ ), edge coordination ( $\eta^2$ ,  $\text{C}_{2v}$ ) and vertex coordination ( $\eta^1$ ,  $\text{C}_{3v}$ ), show that only the face- and edge-coordinated isomers are stable, with the face-coordinated isomer being slightly more stable than the edge-coordinated isomer (2.0  $\text{kJ mol}^{-1}$  at CCSD(T) and 3.5  $\text{kJ mol}^{-1}$  at CASPT2). Therefore, both isomers may coexist under the experimental conditions. As a result of symmetry reduction by argon coordination, the triple degeneracy of the antisymmetric iridium oxide stretching modes is lifted, and the  $T_2$  mode splits into distinct modes. Calculations at the B3LYP and CCSD(T) levels show very small mode splitting (0.2  $\text{cm}^{-1}$  at CCSD(T), 1.5  $\text{cm}^{-1}$  at B3LYP-D3 and 2.0  $\text{cm}^{-1}$  at SO-DFT/B3LYP for the face-coordinated isomer) because the predicted O...Ar distances are quite large and the  $[\text{IrO}_4]^+$  moiety in  $[\text{IrO}_4]^+ \cdot \text{Ar}$  has essentially the same structure as the free cation. Additional *ab initio* multi-reference-based CASPT2 calculations found that there are multi-reference features in  $[\text{IrO}_4]^+$  and  $[\text{IrO}_4]^+ \cdot \text{Ar}$ . The optimized structures (Supplementary Information) show that the O...Ar distances are reduced by comparison

with single-reference methods and that the  $\text{IrO}_4$  fragment exhibits some structural distortion from tetrahedral symmetry, which will surely cause a larger mode splitting than is calculated at the B3LYP and CCSD(T) levels of theory.

Further evidence for the above assignment is gained from additional argon atom coordination. The infrared spectra of the  $[\text{IrO}_4]^+ \cdot \text{Ar}_n$  cations with  $n = 2-4$  are shown in Fig. 2b–d. Owing to symmetry reduction, further mode splitting is observed in the case of  $[\text{IrO}_4]^+ \cdot \text{Ar}_2$ , where the lower-wavenumber band is broad, suggesting the presence of unresolved bands. The spectrum of  $[\text{IrO}_4]^+ \cdot \text{Ar}_3$  shown in Fig. 2c is about the same as that of the single-argon-coordinated complex because the  $[\text{IrO}_4]^+$  cation of the  $[\text{IrO}_4]^+ \cdot \text{Ar}_3$  complex can retain the same symmetry as the  $[\text{IrO}_4]^+$  cation in  $[\text{IrO}_4]^+ \cdot \text{Ar}$ . If the experimental conditions are varied, a spectrum involving three distinct bands at 928, 938 and  $944 \text{ cm}^{-1}$  can clearly be resolved for the  $[\text{IrO}_4]^+ \cdot \text{Ar}_3$  cation, suggesting that additional isomers with mixed coordination modes can be formed (Supplementary Information). When four argon atoms were tagged, only one band at  $939 \text{ cm}^{-1}$  was observed in the spectrum of the  $[\text{IrO}_4]^+ \cdot \text{Ar}_4$  cation (Fig. 2d). In this complex, the tetrahedral symmetry of  $[\text{IrO}_4]^+$  is retained, and so no splitting due to symmetry reduction is expected.

Apart from the bands assigned to the iridium tetroxide cation, three additional absorptions in the experimental spectrum of  $[\text{IrO}_4]^+ \cdot \text{Ar}$  were attributed to different vibrational modes of the  $[(\eta^2\text{-O}_2)\text{IrO}_2]^+$  complex (Table 1 and Fig. 3). The assignment of the  $1,054 \text{ cm}^{-1}$  band to the O–O stretching vibration is consistent with the identification of  $[(\eta^2\text{-O}_2)\text{IrO}_2]^+$  as a cationic peroxide complex<sup>20</sup>. The unprecedented Ir(VII) oxidation state of  $[(\eta^2\text{-O}_2)\text{IrO}_2]^+$  now closes the gap between the well-known VI and the recently discovered VIII<sup>9</sup> oxidation states of this metal. Further evidence for the assignment of the vibrational bands is provided by comparison with the isoelectronic compounds  $\text{OsO}_4$  and  $[(\eta^2\text{-O}_2)\text{OsO}_2]$  previously investigated by matrix-isolation spectroscopy<sup>21</sup> (Supplementary Information). Although this  $[(\eta^2\text{-O}_2)\text{IrO}_2]^+$  complex was predicted at CCSD(T) level to be  $89 \text{ kJ mol}^{-1}$  higher in energy than the tetroxide cation isomer, both structures were experimentally observed, most probably due to the relatively high barrier for the interconversion between the two isomers (see above). With both the  $[(\eta^2\text{-O}_2)\text{Ir}^{\text{VII}}\text{O}_2]^+$  peroxide complex and the  $[\text{Ir}^{\text{IX}}\text{O}_4]^+$  tetroxide complex experimentally characterized here, all possible positive oxidation states of iridium ranging from I to IX are now known.

As well as this gas-phase spectroscopic characterization, experimental attempts have been undertaken to isolate iridium compounds in the IX oxidation state. Proposed syntheses of stable  $[\text{IrO}_4]^+$  salts by the reaction of  $[\text{O}_2][\text{SbF}_6]$  or  $[\text{O}_2][\text{Al}(\text{OC}(\text{CF}_3)_3)_4]$  with  $\text{IrO}_2$  were motivated by the calculated Gibbs free energies ( $\Delta G^\circ$ ) derived from the appropriate Born–Haber cycles: the reactions forming  $[\text{IrO}_4][\text{SbF}_6]$  and  $[\text{IrO}_4][\text{Al}(\text{OC}(\text{CF}_3)_3)_4]$  were predicted to be exergonic by  $-14$  and  $-66 \text{ kJ mol}^{-1}$ , respectively<sup>10</sup>. The use of the weakly coordinating anion  $[\text{Sb}_2\text{F}_{11}]^-$  would also be expected to lead to an exergonic reaction having a  $\Delta G^\circ$  value that is intermediate with respect to  $[\text{SbF}_6]^-$  and  $[\text{Al}(\text{OC}(\text{CF}_3)_3)_4]^-$ . Considering the possibility of a thermally unstable product, the reactions of  $[\text{O}_2][\text{Sb}_2\text{F}_{11}]$  and  $\text{IrO}_2$  were initially attempted at low temperatures ( $-120$  to  $-78^\circ \text{C}$ ) in  $\text{SO}_2\text{ClF}$  and anhydrous HF solvents. The  $\text{SO}_2\text{ClF}$  solution instantly turned bright purple, which was shown to result from the interaction between  $\text{O}_2^+$  and  $\text{SO}_2\text{ClF}$ ; that is, dissolution of  $[\text{O}_2][\text{Sb}_2\text{F}_{11}]$  in  $\text{SO}_2\text{ClF}$  under similar conditions also yielded a bright purple solution. Attempts to synthesize a stable  $[\text{IrO}_4][\text{Sb}_n\text{F}_{5n+1}]$  salt by the reaction of  $[\text{O}_2][\text{Sb}_2\text{F}_{11}]$  and  $\text{IrO}_2$  in superacidic mixtures of HF and  $\text{SbF}_5$  also gave purple solutions, but no evidence for  $[\text{IrO}_4]^+$  was found. The purple solutions are probably attributable to polyoxygen fluoride radicals,  $(\text{O}_2)_{n+1}\text{F}$ , as previously described<sup>22</sup>. To further test the oxidizability of  $\text{IrO}_2$ , the reaction of molten  $\text{XeF}_6$  with  $\text{IrO}_2$  was attempted to determine whether an iridium oxide fluoride species in a higher oxidation state could be formed. There was no apparent reaction even after heating the reaction mixture to  $100^\circ \text{C}$  for over one hour. However, the addition of anhydrous HF to the aforementioned reaction mixture of  $\text{IrO}_2$  and  $\text{XeF}_6$  afforded a dark brown solution at  $21^\circ \text{C}$ , from which crystals of  $[\text{Xe}_2\text{F}_{11}][\text{IrF}_6]$

were grown. The crystal structure of  $[\text{Xe}_2\text{F}_{11}][\text{IrF}_6]$ , and related work, will be reported elsewhere. Although  $\text{IrO}_2$  reacted with  $\text{XeF}_6$ , the high lattice enthalpy associated with the rutile structure of  $\text{IrO}_2$  may significantly inhibit its reactivity with  $\text{O}_2^+$ . Details of the aforementioned attempts to synthesize an  $[\text{IrO}_4]^+$  salt are provided in Supplementary Information.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 1 May 2013; accepted 21 August 2014.

- Jørgensen, C. K. *Oxidation Numbers and Oxidation States* (Springer, 1969).
- Riedel, S. & Kaupp, M. The highest oxidation states of the transition metal elements. *Coord. Chem. Rev.* **253**, 606–624 (2009).
- Riedel, S. in *Comprehensive Inorganic Chemistry II* (eds Reedijk, J. & Poeppelemeier, K.) 187–221 (Elsevier, 2013).
- Schlöder, T. & Riedel, S. in *Comprehensive Inorganic Chemistry* Vol. 9 (ed. Alvarez, S.) 227–243 (Elsevier, 2013).
- Jørgensen, C. K. New understanding of unusual oxidation states in the transition groups. *Naturwissenschaften* **63**, 292 (1976).
- Pyykkö, P., Runeberg, N., Straka, M. & Dylla, K. G. Could uranium(XII) hexoxide,  $\text{UO}_6(\text{O})_6$  exist? *Chem. Phys. Lett.* **328**, 415–419 (2000).
- Xiao, H., Hu, H.-S., Schwarz, W. H. E. & Li, J. Theoretical investigations of geometry, electronic structure and stability of  $\text{UO}_6$ : octahedral uranium hexoxide and its isomers. *J. Phys. Chem. A* **114**, 8837–8844 (2010).
- Gerken, M. & Schrobilgen, G. J. Solution multi-NMR and Raman spectroscopic studies of thermodynamically unstable  $\text{XeO}_4$ . The first  $^{131}\text{Xe}$  NMR study of a chemically bound xenon species. *Inorg. Chem.* **41**, 198–204 (2002).
- Gong, Y., Zhou, M., Kaupp, M. & Riedel, S. Formation and characterization of the iridium tetroxide molecule with iridium in the oxidation state VIII. *Angew. Chem. Int. Ed.* **48**, 7879–7883 (2009).
- Himmel, D., Knapp, C., Patzschke, M. & Riedel, S. How far can we go? Quantum-chemical investigations of oxidation state IX. *ChemPhysChem* **11**, 865–869 (2010).
- Rother, P., Wagner, F. & Zahn, U. Chemical consequences of the  $^{193}\text{Os}(\beta^-)^{193}\text{Ir}$  decay in osmium compounds studied by the Mössbauer method. *Radiochim. Acta* **11**, 203–210 (1969).
- Koyanagi, G. K., Caraiman, D., Blagojevic, V. & Bohme, D. K. Gas-phase reactions of transition-metal ions with molecular oxygen: room-temperature kinetics and periodicities in reactivity. *J. Phys. Chem. A* **106**, 4581–4590 (2002).
- Wang, G. et al. Infrared photodissociation spectroscopy of mononuclear iron carbonyl anions. *J. Phys. Chem. A* **116**, 2484–2489 (2012).
- Duncan, M. A. Infrared spectroscopy to probe structure and dynamics in metal ion-molecule complexes. *Int. Rev. Phys. Chem.* **22**, 407–435 (2003).
- Okumura, M., Yeh, L. I., Myers, J. D. & Lee, Y. T. Infrared spectra of the solvated hydronium ion: vibrational predissociation spectroscopy of mass-selected  $\text{H}_3\text{O}^+(\text{H}_2\text{O})_n(\text{H}_2)_m$ . *J. Phys. Chem.* **94**, 3416–3427 (1990).
- Bieske, E. J. & Dopfer, O. High-resolution spectroscopy of cluster ions. *Chem. Rev.* **100**, 3963–3998 (2000).
- Robertson, W. H. & Johnson, M. A. Molecular aspects of halide ion hydration: the cluster approach. *Annu. Rev. Phys. Chem.* **54**, 173–213 (2003).
- Bach, R. D., Ayala, P. Y. & Schlegel, H. B. A reassessment of the bond dissociation energies of peroxides. An ab initio study. *J. Am. Chem. Soc.* **118**, 12758–12765 (1996).
- Armentrout, P. B. & Li, F. X. Bond energy of  $\text{IrO}^+$ : guided ion-beam and theoretical studies of the reaction of  $\text{Ir}^+(^6\text{F})$  with  $\text{O}_2$ . *J. Phys. Chem. A* **117**, 7754–7766 (2013).
- Gong, Y., Zhou, M. F. & Andrews, L. Spectroscopic and theoretical studies of transition metal oxides and dioxygen complexes. *Chem. Rev.* **109**, 6765–6808 (2009).
- Zhou, M. F., Citra, A., Liang, B. Y. & Andrews, L. Infrared spectra and density functional calculations for  $\text{MO}_2$ ,  $\text{MO}_3$ ,  $(\text{O}_2)\text{MO}_2$ ,  $\text{MO}_4$ ,  $\text{MO}_5^-$  ( $\text{M} = \text{Re}, \text{Ru}, \text{Os}$ ) and  $\text{ReO}_3$ ,  $\text{ReO}_4^-$  in solid neon and argon. *J. Phys. Chem. A* **104**, 3457–3465 (2000).
- Christe, K. O., Wilson, R. D. & Goldberg, I. B. Some observations on the reaction chemistry of dioxygenyl salts and on the blue and purple compounds believed to be  $\text{ClF}_3\text{O}_2$ . *J. Fluor. Chem.* **7**, 543–549 (1976).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** This work was supported by the Ministry of Science and Technology of China (2013CB834603 and 2012YQ220113-3), the National Natural Science Foundation of China (grant nos 21173053, 21433005 and 91026003), the Committee of Science and Technology of Shanghai (13XD1400800), the Fonds der Chemischen Industrie and the GRK 1582 ‘Fluorine as a key element’. We also acknowledge the Natural Sciences and Engineering Research Council of Canada for a Discovery Grant (G.J.S.) and for a postgraduate scholarship (J.T.G.). We are grateful to I. Krossing and H. Hillebrecht for their support.

**Author Contributions** G.W. and M.Z. designed and performed the gas-phase experiments. J.T.G. and G.J.S. attempted to synthesize  $[\text{IrO}_4]$  salts. J.S., J.L., T.S. and S.R. performed the quantum chemical calculations. M.Z., G.J.S., J.L. and S.R. wrote the paper and supervised the experimental and theoretical parts. All authors discussed the results and commented on the manuscript at all stages.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.Z. (mfzhou@fudan.edu.cn), G.J.S. (schrobil@mcmaster.ca) or S.R. (riedel@psichem.de).



## METHODS

**Experimental details.** Iridium oxide cations were generated in the gas phase using a pulsed-laser vaporization/supersonic expansion source. Reactive iridium atoms and cations were generated by the 1,064-nm fundamental of a Nd:YAG laser that was focused onto a rotating iridium metal target. Helium or argon gas, seeded with a few per cent of O<sub>2</sub>, was expanded out of a pulsed valve (General Valve, Series 9) at a stagnation pressure of 8–13 atm. The iridium oxide ions and their argon-tagged complexes generated by laser vaporization were collisionally cooled in the expansion. After free expansion, the cations were skimmed into a second differentially pumped chamber, where the cations were pulse-extracted into a collinear tandem time-of-flight mass spectrometer. Ions were mass-selected by their flight times and then studied by infrared laser photodissociation spectroscopy using a tunable infrared optical parametric oscillator/optical parametric amplifier laser system (LaserVision, pumped by a Continuum Powerlite 8000 Nd:YAG laser). The photodissociation spectrum was obtained by monitoring the yields of the fragment ions as a function of the dissociation infrared laser wavelength and normalizing to the parent ion signal.

**Theoretical and computational details.** The iridium oxides and argon-tagged species were studied by using both density functional theory (DFT) and wavefunction-based theory (WFT). The structures of all molecules were fully optimized by relaxing all geometric parameters at the DFT level using the B3LYP functional<sup>23–26</sup>. The choice of this functional was based on its good performance for comparable molecules<sup>27</sup>. To determine the ground states of the molecules, several spin multiplicities were calculated. In the case of the weak argon complexes, an empirical correction for dispersive interactions (D3) was included in the DFT calculations<sup>28</sup>. Further symmetrical stationary points on the potential energy surface were calculated within the restrictions of the given point groups. Dunning's correlation-consistent triple- $\zeta$  basis sets (aug-cc-pVTZ) were used for both oxygen and argon atoms<sup>29,30</sup>. Scalar relativistic effects were considered by using relativistic energy-consistent small-core pseudopotentials for the metal atom and the corresponding aug-cc-pVTZ-PP basis sets<sup>31,32</sup>. For brevity, this basis set combination is referred to as aug-cc-pVTZ(-PP). In the DFT calculations, averaged masses were used for all elements if not otherwise noted. All standard DFT calculations were performed with the Gaussian 09 program package<sup>23</sup>, whereas the Turbomole V6.4 suite of programs was used for DFT-D3 calculations<sup>33</sup>.

Beyond these calculations, the influence of spin–orbit coupling has been investigated for the energy-minimized structures as well as for the energy barriers. To estimate the interstate crossing energy barrier of the transition from the  $[\text{IrO}_2(\eta^2\text{-O}_2)]^+$  ( $d^2$ ) singlet to the  $[\text{IrO}_4]^+$  ( $d^0$ ) singlet, a series of linear transit calculations were performed by using the DFT/B3LYP method. The starting point of the linear transit energy curves corresponds to the optimized  $[\text{IrO}_2(\eta^2\text{-O}_2)]^+$  structure and the ending point to the optimized  $[\text{IrO}_4]^+$  structure. The linear transit total energies were obtained by constrained optimizations at each linear transit coordinate along the distance from Ir to the centre of the  $\eta^2\text{-O}_2$  configuration. The Slater basis sets with the quality of triple- $\zeta$  plus two polarization functions<sup>34</sup> were used, with the frozen-core approximation applied to the inner shells  $[1s^2-4f^{14}]$  for Ir and  $[1s^2]$  for O. In these calculations, the scalar relativistic and spin–orbit effects were taken into account by the zeroth-order regular approximation<sup>35</sup>. With inclusion of scalar relativistic effects, both the singlet and triplet potential energy surfaces were explored. To further investigate the spin–orbit effects, unrestricted Kohn–Sham calculations with a non-collinear spin–orbit approach were also used in the linear transit calculations. These linear transit calculations were done with Amsterdam density functional (ADF 2013.01) program<sup>36–38</sup>.

Further *ab initio* WFT calculations were performed using coupled-cluster theory at the CCSD(T) level, which used the B3LYP structures with retention of the molecular symmetries. In the CCSD(T) calculations, spin-restricted open-shell Hartree–Fock reference wavefunctions were used in the case of open-shell electron configurations. The CCSD(T) calculations were done using the frozen-core approximation with the  $1s^2$  (O),  $1s^2 2s^2 2p^6$  (Ar) and  $5s^2 5p^6$  (Ir) orbitals excluded from the evaluation of the correlation energies. Stationary points on the potential energy surface were characterized by harmonic vibrational frequency calculations for both the <sup>16</sup>O and the <sup>18</sup>O isotopomers. For all other elements, the mass of the most abundant isotope was used in the CCSD(T) calculations. *Ab initio* coupled-cluster calculations were done using the CFOUR program package<sup>39</sup>.

To better account for the non-dynamic electron correlation arising from electronic transitions from O  $2p$  lone pairs to Ir  $5d$  orbitals, we applied *ab initio* complete active space second-order perturbation theory (CASPT2), as implemented in MOLPRO 2008.1<sup>40</sup>, to do ground-state geometry optimizations of the  $[\text{IrO}_4]^+$  and  $[\text{IrO}_4]^+ \cdot \text{Ar}$  cations using the same aug-cc-pVTZ(-PP) basis sets. In the CASPT2

calculations of  $[\text{IrO}_4]^+$  with  $T_d$  symmetry, the active space contains the lowest five virtually occupied orbitals (that is, Ir–O antibonding orbitals of  $e$  and  $t_2$  symmetry) and the highest six occupied orbitals of  $t_1$  and  $t_2$  symmetry (that is, non-bonding orbitals of O  $2p$  character), which gives 12 electrons in 11 orbitals, that is, CAS(12, 11). Similar active spaces were chosen for  $[\text{IrO}_4]^+ \cdot \text{Ar}$  isomers of  $C_{3v}$ ,  $C_{2v}$  and  $C_s$  symmetry, respectively. A larger active space containing the lowest four unoccupied orbitals and all the occupied valence orbitals except the Ir non-bonding  $5d$  orbital of  $a_1$  symmetry and the O–O  $\sigma$ -bonding orbital of  $a_1$  symmetry, which includes 20 electrons and 14 orbitals, that is, CAS(20, 14) was chosen for the  $[(\text{O}_2)\text{IrO}_2]^+$  cation with  $C_{2v}$  symmetry. In CASPT2 calculations, the  $1s^2$  (O),  $1s^2 2s^2 2p^6$  (Ar) and  $5s^2 5p^6$  (Ir) orbitals were not correlated, to save computing time.

**Assignment of oxidation states.** Usually the formal oxidation state of a central atom in a coordination sphere is defined as the charge of the central atom when every ligand of the coordination sphere is removed in its most stable form. The bonding electron pairs between the metal centre and the ligands are therefore exclusively assigned to the more electronegative fragment, resulting in negative oxidation states. For instance, for fluorine and doubly bonded oxygen ligands the oxidation states of fluorine and oxygen are –I and –II, respectively. According to these rules, the oxidation number of iridium in the  $[\text{IrO}_4]^+$  cation is IX, whereas for the isoelectronic OsO<sub>4</sub> molecule, an oxidation state of VIII is obtained for osmium. However, the concept of oxidation states is a formal one and it has to be kept in mind that calculated atomic charges cannot be used to directly determine oxidation numbers<sup>4</sup>. The former mostly depend on the electronegativities of the ligands and not on the oxidation state of the metal, as exemplified by the isoelectronic  $d^0$  anions  $[\text{V}^{\text{V}}\text{O}_4]^{3-}$ ,  $[\text{Cr}^{\text{VI}}\text{O}_4]^{2-}$  and  $[\text{Mn}^{\text{VII}}\text{O}_4]^-$ , which all represent the highest possible oxidation states of the metal elements, namely V, VI and VII, and for which almost identical charges of 1.05, 1.09 and 0.96 were computed for the central atoms<sup>44,41</sup>. We note that the computed charge of Cr(VI) is, in fact, higher than that of Mn(VII) (refs 4, 41). The calculated atomic charges at the B3LYP/aug-cc-pVTZ(-PP) level using natural population analyses of osmium and iridium in OsO<sub>4</sub> and  $[\text{IrO}_4]^+$  are 1.475 and 1.470, respectively. These atomic charges amount to only a small fraction of the formal oxidation state, which is, however, well established in the case of Os<sup>VIII</sup>O<sub>4</sub>, and there is thus no doubt about the assignment of the formal oxidation state of  $[\text{Ir}^{\text{IX}}\text{O}_4]^+$ , even if the calculated charge of Ir(IX) is slightly less than that of Os(VIII).

23. Frisch, M. J. *et al.* Gaussian09, revision A.1 (Gaussian, Inc., 2009).
24. Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **38**, 3098–3100 (1988).
25. Becke, A. D. Density functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **98**, 5648–5652 (1993).
26. Lee, C., Yang, W. & Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **37**, 785–789 (1988).
27. Riedel, S., Straka, M. & Kaupp, M. Validation of density functional methods for computing structures and energies of mercury(IV) complexes. *Phys. Chem. Chem. Phys.* **6**, 1122–1127 (2004).
28. Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. A consistent and accurate *ab initio* parametrization of density functional dispersion correction (DFT-D) for the 94 elements H–Pu. *J. Chem. Phys.* **132**, 154104 (2010).
29. Kendall, R. A., Dunning, T. H. & Harrison, R. J. Jr. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.* **96**, 6796–6806 (1992).
30. Woon, D. E. & Dunning, T. H. Jr. Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon. *J. Chem. Phys.* **98**, 1358–1371 (1993).
31. Peterson, K. A., Figgen, D., Dolg, M. & Stoll, H. Energy-consistent relativistic pseudopotentials and correlation consistent basis sets for the 4d elements Y–Pd. *J. Chem. Phys.* **126**, 124101 (2007).
32. Figgen, D., Peterson, K. A., Dolg, M. & Stoll, H. Energy-consistent pseudopotentials and correlation consistent basis sets for the 5d elements Hf–Pt. *J. Chem. Phys.* **130**, 164108 (2009).
33. Turbomole Version 6.2, <http://www.turbomole.com> (TURBOMOLE GmbH, 2011).
34. Van Lenthe, E. & Baerends, E. J. Optimized Slater-type basis sets for the elements 1–118. *J. Comput. Chem.* **24**, 1142–1156 (2003).
35. Van Lenthe, E., Baerends, E. J. & Snijders, J. G. Relativistic regular two-component Hamiltonians. *J. Chem. Phys.* **99**, 4597–4610 (1993).
36. ADF v 2013.01, <http://www.scm.com> (SCM, 2013).
37. Fonseca Guerra, C., Snijders, J. G., Te Velde, G. & Baerends, E. J. Towards an order-N DFT method. *Theor. Chem. Acc.* **99**, 391–403 (1998).
38. te Velde, G. *et al.* Chemistry with ADF. *J. Comput. Chem.* **22**, 931–967 (2001).
39. Stanton, J. F. *et al.* CFOUR 1.2 ed., <http://www.cfour.de> (2010).
40. Werner, H.-J. MOLPRO version 2008.1, <http://www.molpro.net> (2008).
41. Aullón, G. & Alvarez, S. Oxidation states, atomic charges and orbital populations in transition metal complexes. *Theor. Chem. Acc.* **123**, 67–73 (2009).

# Methane dynamics regulated by microbial community response to permafrost thaw

Carmody K. McCalley<sup>1†</sup>, Ben J. Woodcroft<sup>2</sup>, Suzanne B. Hodgkins<sup>3</sup>, Richard A. Wehr<sup>1</sup>, Eun-Hae Kim<sup>4</sup>, Rhiannon Mondav<sup>2†</sup>, Patrick M. Crill<sup>5</sup>, Jeffrey P. Chanton<sup>3</sup>, Virginia I. Rich<sup>4</sup>, Gene W. Tyson<sup>2</sup> & Scott R. Saleska<sup>1</sup>

Permafrost contains about 50% of the global soil carbon<sup>1</sup>. It is thought that the thawing of permafrost can lead to a loss of soil carbon in the form of methane and carbon dioxide emissions<sup>2,3</sup>. The magnitude of the resulting positive climate feedback of such greenhouse gas emissions is still unknown<sup>3</sup> and may to a large extent depend on the poorly understood role of microbial community composition in regulating the metabolic processes that drive such ecosystem-scale greenhouse gas fluxes. Here we show that changes in vegetation and increasing methane emissions with permafrost thaw are associated with a switch from hydrogenotrophic to partly acetoclastic methanogenesis, resulting in a large shift in the  $\delta^{13}\text{C}$  signature (10–15‰) of emitted methane. We used a natural landscape gradient of permafrost thaw in northern Sweden<sup>4,5</sup> as a model to investigate the role of microbial communities in regulating methane cycling, and to test whether a knowledge of community dynamics could improve predictions of carbon emissions under loss of permafrost. Abundance of the methanogen *Candidatus 'Methanoflorens stordalenmirensis'*<sup>6</sup> is a key predictor of the shifts in methane isotopes, which in turn predicts the proportions of carbon emitted as methane and as carbon dioxide, an important factor for simulating the climate feedback associated with permafrost thaw in global models<sup>3,7</sup>. By showing that the abundance of key microbial lineages can be used to predict atmospherically relevant patterns in methane isotopes and the proportion of carbon metabolized to methane during permafrost thaw, we establish a basis for scaling changing microbial communities to ecosystem isotope dynamics. Our findings indicate that microbial ecology may be important in ecosystem-scale responses to global change.

Multiple factors—including hydrology, vegetation, organic matter chemistry, pH and soil microclimate—are affected by permafrost loss<sup>5,8,9</sup>. Together these factors regulate microbial metabolisms that release carbon dioxide ( $\text{CO}_2$ ) and methane ( $\text{CH}_4$ ) from thawing permafrost<sup>10–12</sup> and are the basis for Earth-system model predictions of future  $\text{CH}_4$  emissions<sup>7,13,14</sup>. However, the role of microbial community composition in regulating the metabolic processes that drive ecosystem-scale fluxes is unknown.

At our study site in Stordalen mire, as in other thawing permafrost peatlands<sup>8,15</sup>, permafrost loss causes hydrological and vegetation shifts: well-drained permafrost-supported palsas collapse into partly thawed bogs dominated by moss (*Sphagnum* spp.) and fully thawed fens dominated by sedges (such as *Eriophorum angustifolium*)<sup>4</sup>. Between 1970 and 2000, 10% of Stordalen's palsa habitat thawed into such wetlands<sup>4</sup>. This transition drives an appreciable global warming impact because  $\text{CO}_2$ -emitting palsa is converted to bogs and fens, which take up  $\text{CO}_2$  but emit  $\text{CH}_4$  (a more potent greenhouse gas)<sup>3,4,5,16</sup>. The net effect is that the high-methane-emitting fen contributes sevenfold as much greenhouse impact per unit area as the palsa. This thaw progression is also associated with an increase in overall organic matter lability, including a decrease in C:N ratio and an increase in humification rates<sup>9</sup>. We speculated, consistent

with previous studies of *in situ* bog and fen systems<sup>17–19</sup>, that thaw progression also facilitates a shift from hydrogenotrophic to acetoclastic  $\text{CH}_4$  production.

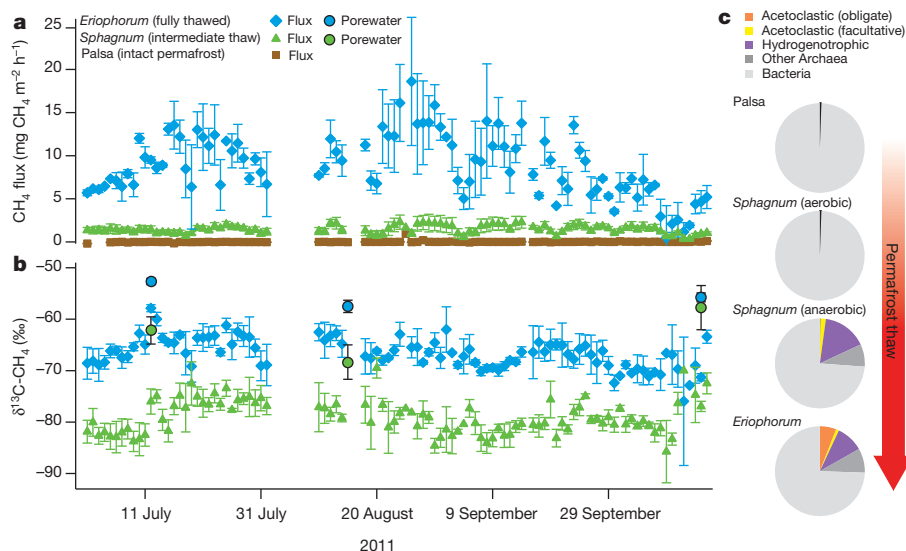
We used the distinct isotopic signatures of different microbial  $\text{CH}_4$  production and consumption pathways to directly relate changes in  $\text{CH}_4$  dynamics across the thaw gradient to underlying changes in the microbial community. Methane produced by hydrogenotrophic methanogens generally has lower  $\delta^{13}\text{C}$  and higher  $\delta\text{D}$  ( $\delta^{13}\text{C} = -110\text{‰}$  to  $-60\text{‰}$  and  $\delta\text{D} = -250\text{‰}$  to  $-170\text{‰}$ ) relative to that produced by acetoclastic methanogens ( $\delta^{13}\text{C} = -60\text{‰}$  to  $-50\text{‰}$  and  $\delta\text{D} = -400\text{‰}$  to  $-250\text{‰}$ )<sup>19,20</sup>. If methanotrophic microbes then oxidize  $\text{CH}_4$ , lighter molecules are preferentially consumed, leaving the remaining  $\text{CH}_4$  enriched in  $^{13}\text{C}$  and D relative to the original  $\text{CH}_4$  pool (expected patterns are shown in Extended Data Fig. 1)<sup>19,20</sup>.

High-temporal-resolution measurements of the magnitude and isotopic composition of  $\text{CH}_4$  emissions, using a quantum cascade laser spectrometer (Aerodyne Research Inc.) connected to autochambers, showed that  $\text{CH}_4$  emissions and their  $^{13}\text{C}$  content increased with thaw. Average  $\text{CH}_4$  fluxes increased from effectively zero at the intact permafrost palsa site to  $1.46 \pm 0.37 \text{ mg CH}_4 \text{ m}^{-2} \text{ h}^{-1}$  (all errors are reported as s.e.m.) at the thawing *Sphagnum* site, and to  $8.75 \pm 0.50 \text{ mg CH}_4 \text{ m}^{-2} \text{ h}^{-1}$  at the fully thawed *Eriophorum* site (Fig. 1a;  $P < 0.001$ ). The average  $\delta^{13}\text{C}$  of emitted  $\text{CH}_4$  also increased significantly, from  $-79.6 \pm 0.9\text{‰}$  in the *Sphagnum* site to  $-66.3 \pm 1.6\text{‰}$  in the *Eriophorum* site (Fig. 1b;  $P = 0.03$ ). This consistent 10–15‰ divergence between sites was maintained through the growing season but overlain by parallel fluctuations in  $\delta^{13}\text{C}$ - $\text{CH}_4$ , suggesting that weather patterns exerted a common influence over the magnitude of isotopic fractionation. Porewater  $\text{CH}_4$  isotopes showed a similar pattern, with *Eriophorum* site porewater  $\delta^{13}\text{C}$  about 10‰ higher than that of *Sphagnum* (July and August; Fig. 1b and Extended Data Table 1). Porewater  $\text{CH}_4$  was  $^{13}\text{C}$ -enriched by 5–20‰ relative to emitted  $\text{CH}_4$ , as expected, as a result of diffusive fractionation (Methods, equation (2))<sup>18,21</sup>.

The apparent fractionation factor for carbon in porewater  $\text{CH}_4$  relative to  $\text{CO}_2$ ,  $\alpha_{\text{C}}$  (Methods, equation (2), and Extended Data Table 1), is a related index of changes in  $\text{CH}_4$  production<sup>22</sup>. Greater fractionation is associated with hydrogenotrophic methanogenesis and was found in the thawing *Sphagnum* site ( $\alpha_{\text{C}} = 1.053 \pm 0.002$ ). Significantly less fractionation ( $P = 0.002$ ) associated with more acetoclastic production or with consumption by oxidation was found in the fully thawed *Eriophorum* porewater ( $\alpha_{\text{C}} = 1.046 \pm 0.001$ ). Here, increases in acetoclastic production, not oxidation, best explain isotopic shifts because lower  $\alpha_{\text{C}}$  and higher  $\delta^{13}\text{C}$ - $\text{CH}_4$  are accompanied by significantly lower  $\delta\text{D}$ - $\text{CH}_4$  (Extended Data Fig. 1;  $P < 0.001$ )<sup>19</sup>. This is consistent with the pattern of isotopes in  $\text{CH}_4$  emissions, in incubations of Stordalen peat<sup>9</sup> and in studies showing bog-to-fen shifts from hydrogenotrophic to acetoclastic methanogenesis<sup>17–19</sup>.

The  $\text{CH}_4$  flux and isotope results provide compelling but indirect evidence for changes in  $\text{CH}_4$ -cycling microbial communities with permafrost

<sup>1</sup>Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85721, USA. <sup>2</sup>Australian Centre for Ecogenomics, School of Chemistry and Molecular Biosciences, University of Queensland, Brisbane 4072, Queensland, Australia. <sup>3</sup>Department of Earth, Ocean and Atmospheric Science, Florida State University, Tallahassee, Florida 32306, USA. <sup>4</sup>Department of Soil, Water and Environmental Science, University of Arizona, Tucson, Arizona 85721, USA. <sup>5</sup>Department of Geological Sciences, Stockholm University, Stockholm 106 91, Sweden. <sup>†</sup>Present addresses: Earth Systems Research Center, University of New Hampshire, Durham, New Hampshire 03824, USA (C.K.M.); Department of Ecology and Genetics, Uppsala University, Uppsala 75 236, Sweden (R.M.).



**Figure 1 | Increases in the magnitude and  $\delta^{13}\text{C}$  signature of  $\text{CH}_4$  during permafrost thaw track shifts in methanogenic communities.** **a**, Average daily  $\text{CH}_4$  emissions (error bars represent s.e.m.;  $n = 2\text{--}3$ ). **b**,  $\delta^{13}\text{C}$  composition of emitted and porewater  $\text{CH}_4$  (error bars represent s.e.m.; flux  $n = 2\text{--}3$ , porewater  $n = 6\text{--}9$ ). **c**, Relative abundance of methanogenic groups as inferred

thaw. These microbiological changes could be shifts in activity of particular community members or changes in community composition. We examined the role of community composition through 16S rRNA gene amplicon sequencing. All known methanogens belong to a small number of archaeal lineages within the Euryarchaeota<sup>23</sup>. As expected, the shift from  $\text{CH}_4$ -neutral intact permafrost palsa to  $\text{CH}_4$ -emitting wetland corresponded to a substantial increase in the relative abundance of methanogenic archaeal lineages (Fig. 1c and Extended Data Table 2, 3). In the aerobic palsa and surface *Sphagnum* habitats, methanogens were found in low relative abundance (average less than 0.6%), whereas the anaerobic environments of the *Eriophorum* and deeper (below the water table) *Sphagnum* habitats harboured communities with a substantially higher relative abundance of methanogens (20–30%).

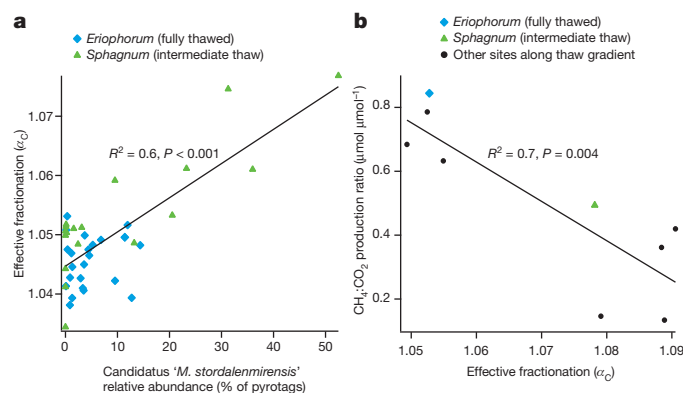
More significantly, the abundance of specific methanogenic lineages varied across the thaw gradient (Fig. 1c and Extended Data Table 2) in a manner corresponding to shifts in  $\text{CH}_4$  production mechanism inferred from the isotope data (Fig. 1b). At the partly thawed *Sphagnum* site, where  $\text{CH}_4$  isotopes were more hydrogenotrophic, the methanogen community was dominated by hydrogenotrophic populations (at least 57% of sequences). Members of the genus *Methanobacterium* and close relatives of the recently described hydrogenotroph Candidatus '*Methanoflorens stordalenmirensis*'<sup>36</sup> (a partial genome of which has also been identified in incubations of Alaskan permafrost<sup>12</sup>) were the most abundant phylogenotypes. Although present, the metabolically versatile *Methanosarcina* (capable of using a wide range of substrates, including acetate and hydrogen<sup>24</sup>), was much less abundant, averaging about 15% of the methanogen sequences. At the fully thawed *Eriophorum* site (where isotope signatures shifted towards acetoclastic), members of the obligately acetoclastic genus *Methanosaeta* increased in abundance, comprising roughly one-third of the methanogenic population. The remaining methanogenic community at the *Eriophorum* site was taxonomically diverse and included lineages also present at the *Sphagnum* site, as well as the hydrogenotrophic genus, *Methanoregula* (Extended Data Table 2). Differences in the functional (hydrogenotrophic versus acetoclastic) composition of the methanogen community between the sites were smallest in October, coinciding with a convergence in  $\delta^{13}\text{C}\text{-CH}_4$  (Fig. 1a and Extended Data Tables 2 and 3).

Taken together, the isotope and microbial sequence data suggest that shifts in microbial communities drive large, concordant variations in  $\text{CH}_4$  isotope biogeochemistry both seasonally and during permafrost

by taxonomic identity assigned from 16S rRNA amplicon sequencing ( $n = 3$ ). For the intermediate-thaw *Sphagnum* site, aerobic communities were sampled above the water table; anaerobic communities were sampled below the water table.

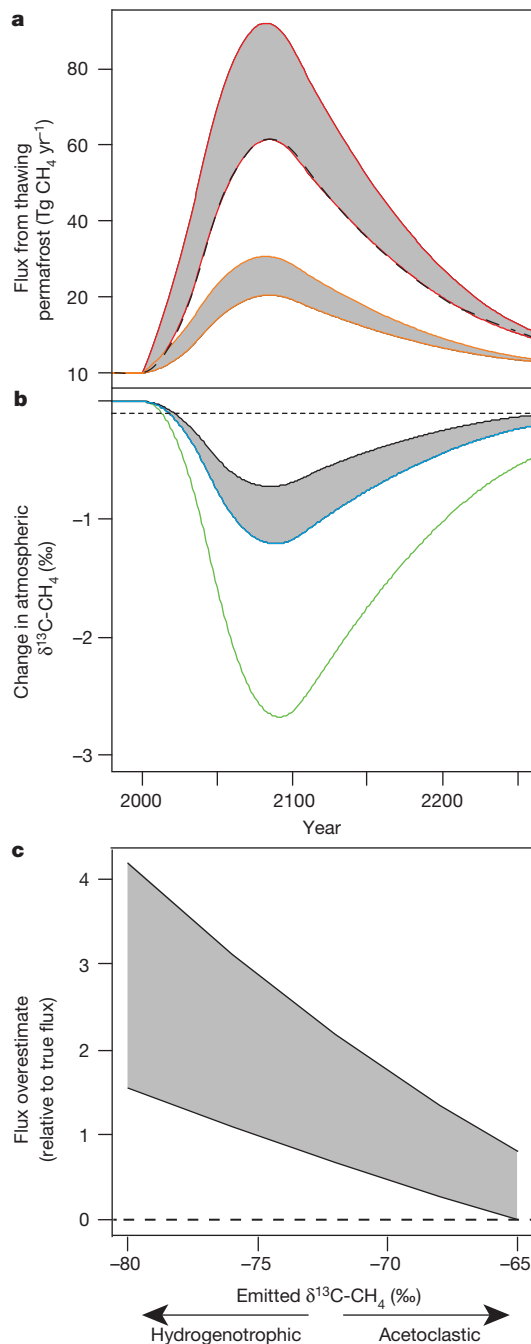
thaw, a novel observation at the ecosystem scale. The early successional hydrogenotroph '*M. stordalenmirensis*'<sup>6</sup> dominates methanogenic metabolism in the early stages of thaw, followed by the subsequent emergence of a more diverse methanogen community, including obligate acetoclastic methanogens. This microbial succession provides direct evidence for how changes in ecosystem structure during permafrost thaw (plant succession and increases in organic matter quality<sup>9</sup>) translate into altered  $\text{CH}_4$  biogeochemistry.

To quantify the effect of this shifting microbial community composition for  $\text{CH}_4$  isotopic patterns, we examined the relationships between isotope fractionation ( $\alpha_c$ ), environmental conditions known or expected to impact methanogenesis, and the relative abundance of specific methanogenic lineages (Extended Data Table 4). Rather than a functional group (such as hydrogenotrophic methanogens), a single organism—the hydrogenotroph '*M. stordalenmirensis*'—was the best one-variable predictor



**Figure 2 | Correlation between  $\alpha_c$  and both Candidatus '*Methanoflorens stordalenmirensis*' and the anaerobic  $\text{CH}_4:\text{CO}_2$  production ratio.** **a**, The relative abundance of a single methanogen, Candidatus '*Methanoflorens stordalenmirensis*', in the field was significantly correlated (linear regression,  $P < 0.001$ ,  $n = 41$ ) with porewater effective fractionation ( $\alpha_c$ ), an isotopic indicator of the methanogenic production pathway. **b**, Anaerobic incubations of peat collected from a related thaw sequence at Stordalen mire (see methods in ref. 9) show a significant correlation between  $\alpha_c$  and the  $\text{CH}_4:\text{CO}_2$  production ratio (linear regression,  $P = 0.004$ ,  $n = 9$ ), suggesting that the abundance of '*M. stordalenmirensis*' may be indicative of the proportion of organic matter metabolized to  $\text{CH}_4$ .





of isotopic patterns in the field (Fig. 2a). Several variables that typically differentiate bogs from fens, including pH and water table depth<sup>18</sup>, were significant predictors of  $\alpha_C$ ; however, it was the relative abundance of '*M. stordalenmirensis*' that explained both the large range of  $\alpha_C$  observed at the *Sphagnum* site ( $R^2 = 0.7$ ,  $P < 0.001$ ) and patterns across sites ( $R^2 = 0.6$ ,  $P < 0.001$ ). This suggests, contrary to the current practice of focusing on the functional diversity of communities, that an individual microbial lineage can have a disproportionate influence on ecosystem biogeochemistry.

Stepwise regression identified environmental variables (water table depth, peat C:N ratio and peat  $\delta^{13}\text{C}$ ) that improved model predictions of  $\alpha_C$  (to  $R^2 = 0.8$ ,  $P < 0.001$ ). Although confirming the central importance of '*M. stordalenmirensis*' in explaining variation in  $\alpha_C$  (Extended Data Table 5), this model also supports the hypothesis that organic matter chemistry underlies shifts in  $\text{CH}_4$  metabolism<sup>9,25</sup>. The dependence on the abundance of this lineage was evident despite the relative rather

**Figure 3 | Simulated effect of  $\text{CH}_4$  from different methanogen communities in thawing permafrost on atmospheric  $\delta^{13}\text{C}-\text{CH}_4$  in a box model of the atmosphere.** **a**, Modelled  $\text{CH}_4$  emissions under high (red bounding lines) and low (orange bounding lines) climate warming scenarios, and a range within each (grey tint) spanning high and low C-release scenarios<sup>2</sup>. **b**, Consequent decreases (simulated by the intermediate emissions scenario indicated by the red dashed line in **a**) in  $\delta^{13}\text{C}$  of atmospheric  $\text{CH}_4$  due to emissions dominated by hydrogenotrophic lineages, as in intermediate-thaw *Sphagnum* sites (green line,  $\delta^{13}\text{C} = -80\text{‰}$ ), or more by acetoclasts, as in fully thawed *Eriophorum* sites (blue line,  $\delta^{13}\text{C} = -65\text{‰}$ ). Atmospheric inversion models typically assume that emissions have  $\delta^{13}\text{C}$  ranging from  $-60\text{‰}$  (black line) to  $-65\text{‰}$  (blue line). (The dotted horizontal line indicates the current detection limit for atmospheric  $\text{CH}_4$  isotopes<sup>28</sup>). These imply an underestimate of the effect on atmospheric  $\delta^{13}\text{C}$  for the given emissions scenario (blue or green). **c**, To match observed atmospheric isotopes, the box model would then require a corresponding overestimate of  $\text{CH}_4$  flux attributed to permafrost thaw (vertical axis). The magnitude of the overestimate depends on the mismatch between model-assumed isotopic composition (upper line,  $-60\text{‰}$ ; lower line,  $-65\text{‰}$ ) and the actual isotopic composition produced by different communities, which ranges here along the horizontal axis from  $-80\text{‰}$  (hydrogenotroph-dominated, as in the partly thawed *Sphagnum* sites) to  $-65\text{‰}$  (acetoclastic, as in the fully thawed *Eriophorum* sites).

than the absolute nature of the community composition analysis, and the measurement of abundance rather than activity. We speculate that direct measures of gene expression or metabolic activity (metatranscriptomics and metaproteomics) will have an even stronger association than community composition data with isotopic signatures.

Further analysis showed that  $\alpha_C$  is significantly correlated ( $R^2 = 0.7$ ,  $P = 0.004$ ) with the large range in  $\text{CH}_4:\text{CO}_2$  production ratio (0.13–0.84) measured in anaerobic incubations of Stordalen peat (Fig. 2b). It is therefore likely that changes in the proportion of anaerobically mineralized C that ends up as  $\text{CH}_4$ —a key, but poorly constrained, parameter in global  $\text{CH}_4$  models<sup>26</sup>—tracks the abundance of '*M. stordalenmirensis*', which acts as an index of the concerted changes in microbial community and organic matter chemistry that together control the efficiency of carbon metabolism.

Incorporating this understanding of the imprint of microbial communities could be crucial to both improved model prediction of future climate change  $\text{CH}_4$  feedbacks and accurate attribution of the portion of global atmospheric  $\text{CH}_4$  change that is derived from permafrost thaw. First, in simulating  $\text{CH}_4$  cycling, Earth-system models typically prescribe as fixed the fraction of anaerobically metabolized carbon that becomes  $\text{CH}_4$  (ref. 26). The lack of a basis for predicting this parameter across ecosystems and in response to climate change limits current modelling efforts<sup>3</sup>. Our finding that the  $\text{CH}_4:\text{CO}_2$  production ratio is highly variable and predictable from isotopic indicators of methanogenic community composition (Fig. 2b) supports the need to improve the representation of microbial ecology in models<sup>17,27</sup>. Although simulating microbial population dynamics is beyond the scope of current global models, the identification of microbial lineages that predict key parameters, such as the  $\text{CH}_4:\text{CO}_2$  ratio, provides insights that improve simulations of  $\text{CH}_4$  biogeochemistry used to estimate global emissions.

Second, atmospheric inversion studies that use  $\text{CH}_4$  mixing ratios and isotopes to infer global sources and sinks of atmospheric  $\text{CH}_4$  assume that wetland microbial sources are dominated by acetate fermentation ( $-58\text{‰}$  to  $-65\text{‰}$ ), and, critically, that isotopic signatures from biological sources are constant over time<sup>28,29</sup>. In contrast, we observed isotopic compositions that varied across a gradient of permafrost thaw: hydrogenotrophic methanogenesis was estimated to produce about 50–75% of total  $\text{CH}_4$  emission at Stordalen (Extended Data Table 6), with  $\delta^{13}\text{C}$  averaging  $-80\text{‰}$  (Fig. 1b). The hydrogenotrophic  $\delta^{13}\text{C}$  observed at Stordalen and other Arctic wetlands<sup>30</sup> may be a ubiquitous characteristic of thawing permafrost, particularly during thaw stages that generate recalcitrant organic matter<sup>9,25</sup>, such as that observed at Stordalen in the intermediate-thaw *Sphagnum* site.

To test whether these observed thaw-induced changes in microbial metabolism might be relevant for large-scale atmospheric methane dynamics,

we used a simple box model of atmospheric mixing (Methods, equation (3)) to quantify the effect of different methanogen communities within recently constructed scenarios of CH<sub>4</sub> emission from thawing permafrost<sup>2</sup> (Extended Data Fig. 2a, b). We found that if hydrogenotrophic lineages regulated CH<sub>4</sub> isotope patterns in permafrost thaw generally, as at Stordalen, then projected CH<sub>4</sub> emissions (Fig. 3a) would produce larger decreases in  $\delta^{13}\text{C}$  of atmospheric CH<sub>4</sub> than expected from current inversion model assumptions that acetoclasts dominate emissions (Fig. 3b and Extended Data Fig. 2c, d). This in turn would constrain our simple box model to substantially overestimate the amount of CH<sub>4</sub> released from thawing permafrost and underestimate emissions from non-wetland sources, most notably fossil fuels (Fig. 3c). The greater the prevalence of hydrogenotrophic lineages in CH<sub>4</sub> emissions, the larger will be the overestimate of fluxes from thaw (Fig. 3c). The numerical size of the mis-estimation error here is illustrative; state-of-the-art three-dimensional inversion models have spatially resolved constraints that would probably force smaller flux mis-estimations. However, the general implication is that microbial effects are sufficiently important that accurate global accounting of the different sources of CH<sub>4</sub> under future climate change can be improved by understanding the microbial community dynamics underlying biological feedbacks in natural systems.

By showing that the abundance of key microbial lineages can be used to predict atmospherically relevant patterns in CH<sub>4</sub> isotopes and the proportion of carbon metabolized to CH<sub>4</sub> during permafrost thaw, this work establishes a basis for scaling changing microbial communities to ecosystem-scale and global-scale atmospheric isotope dynamics. It also highlights the central role of microbial ecology in ecosystem-scale responses to global change and the benefit of incorporating microbial dynamics into Earth-system models.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 19 March; accepted 22 August 2014.**

- Tarnocai, C. *et al.* Soil organic carbon pools in the northern circumpolar permafrost region. *Glob. Biogeochem. Cycles* **23**, GB2023 (2009).
- Schuur, E. A. G. *et al.* Expert assessment of vulnerability of permafrost carbon to climate change. *Clim. Change* **119**, 359–374 (2013).
- Ciais, P. *et al.* *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge University Press, 2013).
- Johansson, T. *et al.* Decadal vegetation changes in a northern peatland, greenhouse gas fluxes and net radiative forcing. *Glob. Change Biol.* **12**, 2352–2369 (2006).
- Christensen, T. R. *et al.* Thawing sub-arctic permafrost: effects on vegetation and methane emissions. *Geophys. Res. Lett.* **31**, L04501 (2004).
- Mondav, R. *et al.* Discovery of a novel methanogen in thawing permafrost. *Nature Commun.* **5**, 3212, <http://dx.doi.org/10.1038/ncomms4212> (14 February 2014).
- Melton, J. R. *et al.* Present state of global wetland extent and wetland methane modelling: conclusions from a model inter-comparison project (WETCHIMP). *Biogeosciences* **10**, 753–788 (2013).
- Jorgenson, M. T., Racine, C. H., Walters, J. C. & Osterkamp, T. E. Permafrost degradation and ecological changes associated with a warming climate in central Alaska. *Clim. Change* **48**, 551–579 (2001).
- Hodgkins, S. B. *et al.* Changes in peat chemistry associated with permafrost thaw increase greenhouse gas production. *Proc. Natl Acad. Sci. USA* **111**, 5819–5824 (2014).
- Olefeldt, D., Turetsky, M. R., Crill, P. M. & McGuire, A. D. Environmental and physical controls on northern terrestrial methane emissions across permafrost zones. *Glob. Change Biol.* **19**, 589–603 (2012).
- Lee, H., Schuur, E. A. G., Inglett, K. S., Lavoie, M. & Chanton, J. P. The rate of permafrost carbon release under aerobic and anaerobic conditions and its potential effects on climate. *Glob. Change Biol.* **18**, 515–527 (2012).
- Mackelprang, R. *et al.* Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw. *Nature* **480**, 368–371 (2011).
- Riley, W. J. *et al.* Barriers to predicting changes in global terrestrial methane fluxes: analyses using CLM4Me, a methane biogeochemistry model integrated in CESM. *Biogeosciences* **8**, 1925–1953 (2011).
- Koven, C. D. *et al.* Permafrost carbon-climate feedbacks accelerate global warming. *Proc. Natl Acad. Sci. USA* **108**, 14769–14774 (2011).
- Turetsky, M. R., Wieder, R. K. & Vitt, D. H. Boreal peatland C fluxes under varying permafrost regimes. *Soil Biol. Biochem.* **34**, 907–912 (2002).
- Bäckstrand, K. *et al.* Annual carbon gas budget for a subarctic peatland, northern Sweden. *Biogeosciences* **7**, 95–108 (2010).
- Bridgman, S. D., Cadillo-Quiroz, H., Keller, J. K. & Zhuang, Q. Methane emissions from wetlands: biogeochemical, microbial, and modeling perspectives from local to global scales. *Glob. Change Biol.* **19**, 1325–1346 (2013).
- Hornibrook, E. R. C. & Bowes, H. L. Trophic status impacts both the magnitude and stable carbon isotope composition of methane flux from peatlands. *Geophys. Res. Lett.* **34**, 2–6 (2007).
- Chanton, J. P., Chaser, L. C., Glaser, P. & Siegel, D. in *Stable Isotopes and Biosphere-Atmosphere Interactions* (eds Flanagan, L. B., Ehleringer, J. R. & Pataki, D. E.) 85–105 (Elsevier, 2005).
- Whiticar, M. J. Carbon and hydrogen isotope systematics of bacterial formation and oxidation of methane. *Chem. Geol.* **161**, 291–314 (1999).
- Popp, T. J., Chanton, J. P., Whiting, G. J. & Grant, N. Methane stable isotope distribution at a Carex dominated fen in North Central Alberta. *Glob. Biogeochem. Cycles* **13**, 1063–1077 (1999).
- Whiticar, M. J., Faber, E. & Schoel, M. Biogenic methane formation in marine and freshwater environments: CO<sub>2</sub> reduction vs. acetate fermentation—isotope evidence. *Geochim. Cosmochim. Acta* **50**, 693–709 (1986).
- Ferry, J. G. How to make a living by exhaling methane. *Annu. Rev. Microbiol.* **64**, 453–473 (2010).
- Liu, Y. & Whitman, W. B. Metabolic, phylogenetic, and ecological diversity of the methanogenic archaea. *Ann. NY Acad. Sci.* **1125**, 171–189 (2008).
- Hornibrook, E. R. C., Longstaffe, F. J. & Fyfe, W. S. Spatial distribution of microbial methane production pathways in temperate zone wetland soils: stable carbon and hydrogen isotope evidence. *Geochim. Cosmochim. Acta* **61**, 745–753 (1997).
- Wania, R. *et al.* Present state of global wetland extent and wetland methane modelling: methodology of a model inter-comparison project (WETCHIMP). *Geoscient. Model. Devel.* **6**, 617–641 (2013).
- Wieder, W. R., Bonan, G. B. & Allison, S. D. Global soil carbon projections are improved by modelling microbial processes. *Nature Clim. Change* **3**, 909–912 (2013).
- Kai, F. M., Tyler, S. C., Randerson, J. T. & Blake, D. R. Reduced methane growth rate explained by decreased Northern Hemisphere microbial sources. *Nature* **476**, 194–197 (2011).
- Bousquet, P. *et al.* Contribution of anthropogenic and natural sources to atmospheric methane variability. *Nature* **443**, 439–443 (2006).
- Hines, M. E., Duddleston, K. N., Rooney-Varga, J. N., Fields, D. & Chanton, J. P. Uncoupling of acetate degradation from methane formation in Alaskan wetlands: connections to vegetation distribution. *Glob. Biogeochem. Cycles* **22**, 1–12 (2008).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank the Abisko Scientific Research Station for infrastructure and logistical support; T. Logan and N. Rakos for their assistance in the field; and S. Wofsy and S. Frolking for feedback on a draft of this paper. This work was supported by the US Department of Energy Office of Biological and Environmental Research (award DE-SC0004632), and by the University of Arizona Technology and Research Initiative Fund, through the Water, Environmental and Energy Solutions Initiative. R.M. was supported by an Australian Postgraduate Award Scholarship.

**Author Contributions** S.R.S., V.I.R., P.M.C., J.C. and G.W.T. designed the study. C.K.M., S.B.H., R.A.W., P.M.C., J.C. and S.R.S. designed and/or performed flux/porewater/isotope measurements and laboratory incubations. C.K.M., B.J.W., R.M., E.-H.K., S.R.S., V.I.R. and G.W.T. designed and/or performed analyses integrating bioinformatics and biogeochemistry. C.K.M., V.I.R. and S.R.S. wrote the paper in consultation with B.J.W., S.B.H., J.C., P.M.C., E.-H.K., R.M. and G.W.T.

**Author Information** Amplicon sequencing data are deposited in the NCBI Sequence Read Archive with accession number SRP042265. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.K.M. ([carlmody.mccalley@unh.edu](mailto:carlmody.mccalley@unh.edu)) or S.R.S. ([saleska@email.arizona.edu](mailto:saleska@email.arizona.edu)).

## METHODS

**Site description and permafrost thaw.** Stordalen is a sub-arctic palsamire located 10 km east of Abisko in the discontinuous permafrost zone of northern Sweden (68° 21' N, 18° 49' E, altitude 363 m above sea level). This work focuses on three distinct subhabitats, common to northern wetlands and together covering about 98% of the mire's surface: permafrost-dominated, well-drained palsas occupied by feather mosses and ericaceous and woody plants, covering 49% of the mire; intermediate permafrost sites with variable water table depth, dominated by *Sphagnum* spp., covering 37% of the mire; and full summer-thaw, wet sites with *Eriophorum angustifolium*, covering 12% of the mire. Between 1970 and 2000, as permafrost thawed and palsas collapsed, *Sphagnum* sites and *Eriophorum* sites expanded by 3% and 54%, respectively<sup>4</sup>.

The formation of wetlands after permafrost thaw, as observed at Stordalen, is a widespread characteristic of peatlands affected by permafrost loss<sup>8,31–33</sup>. Thawing of ice-rich features results in peatland collapse and the formation of bogs and fens. At Stordalen, thaw is associated with a progression from ombrotrophic bogs to minerotrophic fens due to thaw-induced subsidence increasing hydrological connectivity. A similar successional shift from bogs dominated by *Sphagnum* spp. to tall graminoid fens has been observed in other northern peatlands<sup>8,33–35</sup>. More generally, landscape features and hydrological conditions dictate the characteristics and trajectory of wetland communities formed after permafrost thaw<sup>36</sup>. For example, rapid fen development is observed at the subsiding margins of permafrost plateaux<sup>37</sup>, whereas collapse bogs and thermokarst lakes often form within large, thawing peatland complexes<sup>32</sup>. Large uncertainty in model predictions of the extent and characteristics of wetland formation arising from permafrost thaw is a critical limitation to current understanding of carbon–climate feedbacks<sup>7,14</sup>. As demonstrated in this study, improved characterization and modelling of peatland transformation during thaw will be essential for accurately predicting post-thaw microbial communities and the resultant magnitude and isotopic composition of CH<sub>4</sub> emissions under climate change.

**Methane isotope systematics.** We use standard  $\delta$  notation for quantifying the isotopic compositions of CH<sub>4</sub> and CO<sub>2</sub>: the ratio  $R$  of <sup>13</sup>C to <sup>12</sup>C (or D to H) in the measured sample is expressed as a relative difference (denoted  $\delta^{13}\text{C}$  or  $\delta\text{D}$ ) from the Vienna Pee Dee Belemnite (VPDB) international standard material. For example, for C:

$$\delta^{13}\text{C} = \frac{R - R_{\text{VPDB}}}{R_{\text{VPDB}}} = \frac{R}{R_{\text{VPDB}}} - 1 \quad (1)$$

$\delta^{13}\text{C}$  is often expressed in parts per thousand (‰).

Isotopic fractionation in chemical reactions (including methanogenesis or methanotrophy) or due to diffusion may be quantified as

$$\alpha = \frac{R_{\text{source}}}{R_{\text{product}}} = \frac{\delta_{\text{source}} + 1}{\delta_{\text{product}} + 1} \quad (2)$$

For diffusive fractionation,  $R_{\text{source}}$  is taken to be the isotopic ratio in the concentrations of the gradient and  $R_{\text{product}}$  the ratio in the resultant net flux. Because diffusion discriminates against the heavy isotope,  $R_{\text{product}} < R_{\text{source}}$ , which implies, for example, that the isotopic ratio of porewater (the 'source') will be greater than that of the flux of gas diffusing out, as we see here (Fig. 1a). Methanogenesis and methanotrophy also discriminate against the heavier isotopes, so that  $R_{\text{product}} < R_{\text{source}}$  (and hence  $\alpha > 1$ ) for both C and H in methane. Note that  $\alpha > 1$  for methanotrophy implies that the products of CH<sub>4</sub> oxidation (CO<sub>2</sub> and H<sub>2</sub>O) are lighter (have lower  $R$ ) in both C and H relative to the source CH<sub>4</sub>; however, mass balance then requires the residual methane not oxidized to become heavier in both C and H relative to the starting composition of the CH<sub>4</sub> pool before oxidation.

The degree of C isotopic fractionation between CO<sub>2</sub> and CH<sub>4</sub> differs between the two main biochemical pathways of methanogenesis, namely acetoclastic (CH<sub>3</sub>COOH → CH<sub>4</sub> + CO<sub>2</sub>) and hydrogenotrophic (CO<sub>2</sub> + 4H<sub>2</sub> → 2H<sub>2</sub>O + CH<sub>4</sub>). Carbon isotope fractionation ( $\alpha_{\text{C}}$ ) is greater for hydrogenotrophic than for acetoclastic methanogenesis, but  $\alpha_{\text{H}}$  (hydrogen isotope fractionation) follows the opposite pattern:  $\alpha_{\text{H}}$  (hydrogenotrophic) <  $\alpha_{\text{H}}$  (acetoclastic) (Extended Data Fig. 1; ref. 19). Hence, variations in C and H isotopic compositions of CH<sub>4</sub> that arise from variations in methanogenic pathway will be anti-correlated: shifts from hydrogenotrophic to acetoclastic production will cause C isotope ratios to increase but H isotope ratios to decline, moving along a negatively sloped 'production line' in H–C isotope space (Extended Data Fig. 1). Isotopic variations that arise from variations in the degree of methanotrophy, by contrast, will be positively correlated: shifts towards increasing methanotrophy will cause both C and H isotope ratios to increase along a positively sloped 'oxidation line' (Extended Data Fig. 1).

In a field study such as this, it is difficult to estimate fractionation factors directly; we therefore follow standard practice in the methane biogeochemistry literature (see, for example, refs 22, 38) and estimate the net or effective fractionation factor from *in situ* pore water data. For example, we estimate  $\alpha_{\text{C}}$ , the effective fractionation

factor for C in CH<sub>4</sub>, by applying equation (2), setting  $\delta_{\text{product}} = \delta^{13}\text{C}_{\text{CH}_4}$  and  $\delta_{\text{source}} = \delta^{13}\text{C}_{\text{CO}_2}$ , where  $\delta^{13}\text{C}_{\text{CH}_4}$  and  $\delta^{13}\text{C}_{\text{CO}_2}$  are the observed C compositions of CH<sub>4</sub> and CO<sub>2</sub>, respectively<sup>38</sup>. Using CO<sub>2</sub> isotope composition for  $\delta_{\text{source}}$  follows directly for hydrogenotrophic methanogenesis (for which CO<sub>2</sub> is the source C substrate) and has been found to work also in practice for acetoclastic methanogenesis, because porewater CO<sub>2</sub> arises primarily from respiration of organic matter (a non-discriminatory reaction), and so is typically isotopically indistinguishable from organic matter<sup>20,39</sup>.

**Autochamber measurements.** The autochamber system at Stordalen mire has previously been described in detail for measurements of CO<sub>2</sub> and total hydrocarbons<sup>16,40</sup>. In brief, a system of eight automatic gas-sampling chambers made of transparent Lexan was installed in the three habitat types at Stordalen mire in 2001 ( $n = 3$  each in the palsa and *Sphagnum* habitats, and  $n = 2$  in the *Eriophorum* habitat). Each chamber covers an area of 0.14 m<sup>2</sup> (38 cm × 38 cm), with a height of 25–45 cm, and is closed once every 3 h for a period of 5 min. The chambers are connected to the gas analysis system, located in an adjacent temperature controlled cabin, by 3/8-inch Dekoron tubing through which air is circulated at approximately 2.5 l min<sup>−1</sup>. During the 2011 season the system was updated with a new chamber design similar to that described in ref. 41. The new chambers each cover an area of 0.2 m<sup>2</sup> (45 cm × 45 cm), with a height ranging from 15 to 75 cm depending on habitat vegetation. At the Palsa and *Sphagnum* site the chamber base is flush with the ground and the chamber lid (15 cm in height) lifts clear of the base between closures. At the *Eriophorum* site the chamber base is raised 50–60 cm on Lexan skirts to accommodate vegetation of large stature. In addition, each chamber is instrumented with thermocouples measuring air and surface ground temperature, and the depth of the water table is measured manually three to five times per week. The Palsa site chambers are located within the palsa site in ref. 6 and correspond to the hummock site class (I) described in ref. 4. The *Sphagnum* site chambers are located within the bog site in ref. 6 or site S in ref. 9 and correspond to the semi-wet and wet site classes (II and III) described in ref. 4. The *Eriophorum* site chambers are located within the fen site in ref. 6 or site E in ref. 9 and correspond to the tall graminoid site class (IV) described in ref. 4.

**Quantum cascade laser spectrometer measurement and calibration.** Methane fluxes and isotopes were measured with a quantum cascade laser spectrometer (QCLS; Aerodyne Research Inc.), deployed to Stordalen mire in June 2011. The QCLS instrument at Stordalen is a modification of the technology described in detail in ref. 42. In brief, the QCLS uses a room-temperature continuous-wave mid-infrared laser whose frequency was tuned to scan rapidly (900 kHz) across <sup>12</sup>CH<sub>4</sub> and <sup>13</sup>CH<sub>4</sub> absorption lines in the 7.5-μm region. The laser light enters a multipass sample cell (effective path length about 200 m) containing sample air at low pressure (about 5 kPa) and is detected by a thermoelectrically cooled detector (no cryogenics are needed). Aerodyne Research's custom TDL Wintel software averages high-frequency spectra to produce independent <sup>12</sup>CH<sub>4</sub> and <sup>13</sup>CH<sub>4</sub> mixing ratios in the sample airstream at 1-s intervals. The ratio  $R$  of <sup>13</sup>CH<sub>4</sub> to <sup>12</sup>CH<sub>4</sub> can then be expressed in standard notation as  $\delta^{13}\text{C}$ , the part-per-thousand (‰) deviation of the measured ratio from the VPDB standard <sup>13</sup>C/<sup>12</sup>C ratio  $R_{\text{VPDB}}$ , as in equation (1).

Instrument precision in the field at Stordalen mire was assessed by using time-series measurements of calibration tank air over 30–40 min. The precision of  $\delta^{13}\text{C}$ -CH<sub>4</sub> measurements using a 1-s integration time was 1‰. The Allan variance technique (used to characterize the minimum possible measurement error and the averaging time required to achieve it<sup>43</sup>), showed that the minimum measurement error on  $\delta^{13}\text{C}$ -CH<sub>4</sub> was less than 0.2‰, achieved with 60 s of averaging time. This approaches the precision of comparable measurements made with gas chromatography–isotope ratio mass spectrometry (GC–IRMS).

We connected the QCLS to the main autochamber circulation with 1/4-inch Dekoron tubing and a solenoid manifold that enabled selection between the autochamber flow and an array of calibration tanks. During measurement periods, filtered (0.45 μm Teflon filter) and dried (Perma Pure PD-100T-24MSA) sample air flows at 1.4 standard litres per minute through the 2-litre QCLS sample cell volume at 5.6 kPa. A downstream solenoid controls the QCLS return flow so that air recirculates only during autochamber measurement periods; during calibration periods, exhaust air is vented to the room.

Calibrations were performed every 60 min with three calibration gases spanning the observed concentration range (1.5–10 p.p.m.). The CH<sub>4</sub> concentration and  $\delta^{13}\text{C}$  composition of each calibration tank was determined by inter-calibration with a set of four well-characterized primary standard tanks. The primary tanks (Scott Marin, Inc.) were calibrated to the VPDB scale by means of flask samples, which were analysed by GC–IRMS at Florida State University (see porewater methods for GC–IRMS details). Each isotopologue, <sup>12</sup>CH<sub>4</sub> and <sup>13</sup>CH<sub>4</sub>, was treated as an independent measurement and calibrated separately. For each calibration period a linear calibration curve was fitted for each isotopologue and the fit parameters were then linearly interpolated between calibration periods. The interpolated fit parameters were applied



to the measured sample isotopologue mixing ratios to give calibrated measurements of  $^{12}\text{CH}_4$ ,  $^{13}\text{CH}_4$  and total  $\text{CH}_4$ , from which  $\delta^{13}\text{C}-\text{CH}_4$  was calculated.

**Autochamber data processing.** For each autochamber closure we calculated the flux and  $\delta^{13}\text{C}$  signature of emitted  $\text{CH}_4$ . Fluxes were calculated by using a method consistent with that detailed in ref. 44 for  $\text{CO}_2$  and total hydrocarbons, using a linear regression of changing headspace  $\text{CH}_4$  concentration over a period of 2.5 min. Eight 2.5-min regressions were calculated, staggered by 15 s, and the most linear fit (highest  $r^2$ ) was then used to calculate flux. Keeling plots<sup>45–47</sup> using the entire closure period were used to estimate the isotopic composition of the emitted  $\text{CH}_4$ . As demonstrated in ref. 42, negligible error in measurement of  $\text{CH}_4$  relative to that of  $\delta^{13}\text{CH}_4$  for this instrumentation meant that type I regression was sufficient for the Keeling plot analysis. When the total change in headspace  $\text{CH}_4$  was low<sup>45</sup>, there was high error in the Keeling intercept. We used a threshold of 3‰ uncertainty in the Keeling intercept as a cutoff for including isotopic values in the calculation of daily and annual averages, resulting in a total of 1,569 observations at the *Sphagnum* site and 1,168 at the *Eriophorum* site. No Palsa chamber closures had sufficient  $\text{CH}_4$  flux to calculate  $\delta^{13}\text{CH}_4$ . Daily and whole-season average flux and isotopic composition for each habitat were calculated on the basis of individual chambers as the unit of replication ( $n = 3$  for Palsa and *Sphagnum*,  $n = 2$  for *Eriophorum*). Significant differences in the magnitude and isotopic composition of  $\text{CH}_4$  emissions were determined with Student's *t*-test (isotopic composition) and analysis of variance (flux magnitude) in R<sup>48</sup>, with seasonal averages for each autochamber as the unit of replication. Statistical significance was determined at  $\alpha = 0.05$ .

**Porewater sampling and analysis.** Porewater samples were collected on 12 July 2011, 15 August 2011 and 15 October 2011 at three locations adjacent to the *Sphagnum* and *Eriophorum* autochamber sites (Extended Data Table 1). Samples were collected by suction with a syringe through a stainless steel tube and filtered through 25-mm diameter Whatman Grade GF/D glass microfibre filters (2- $\mu\text{m}$  particle retention). Porewater pH was measured in the field (Oakton Waterproof pHTestr 10; Eutech Instruments). Samples for the analysis of the concentration and  $\delta^{13}\text{C}$  of  $\text{CH}_4$  and  $\text{CO}_2$  were injected into 30-ml evacuated vials sealed with butyl rubber septa and frozen within 8 h of collection. The samples for  $\delta\text{D}-\text{CH}_4$  were injected into 120-ml evacuated vials sealed with butyl rubber septa and containing 0.5 g of KOH. For  $\delta\text{D}-\text{H}_2\text{O}$ , water was filtered directly into 2-ml plastic screw-cap vials so that the vials were completely filled, then frozen within 8 h of collection. All samples were shipped frozen to Florida State University for analysis.

Samples collected for analysis of  $\text{CH}_4$  and  $\text{CO}_2$  concentrations and  $\delta^{13}\text{C}$  were thawed, acidified with 0.5 ml of 21%  $\text{H}_3\text{PO}_4$ , and brought to atmospheric pressure with helium. The sample headspace was analysed for concentrations and  $\delta^{13}\text{C}$  of  $\text{CH}_4$  and  $\text{CO}_2$  on a continuous-flow Hewlett-Packard 5890 gas chromatograph (Agilent Technologies) at 40 °C coupled to a Finnigan MAT Delta S isotope ratio mass spectrometer via a ConFlo IV interface system (Thermo Scientific). The headspace gas concentrations were converted to porewater concentrations on the basis of their known extraction efficiencies, defined as the proportion of formerly dissolved gas in the headspace. An extraction efficiency of 0.95 (based on repeated extractions) was used for  $\text{CH}_4$ , and the extraction efficiency for  $\text{CO}_2$  relative to dissolved inorganic carbon (DIC) was determined on the basis of  $\text{CO}_2$  extraction from dissolved bicarbonate standards<sup>49</sup>.

Samples collected for analysis of  $\delta\text{D}-\text{CH}_4$  were brought to atmospheric pressure with helium and measured on a gas chromatograph connected to a ThermoFinnigan Delta Plus continuous-flow isotope ratio mass spectrometer at the National High Magnetic Field Laboratory (Tallahassee, FL).  $\delta\text{D}$  of  $\text{CH}_4$  is affected by  $\delta\text{D}$  of  $\text{H}_2\text{O}$  because  $\text{CH}_4$  exchanges H atoms with water during methanogenesis<sup>20,38,50</sup>, so measurement of  $\delta\text{D}-\text{H}_2\text{O}$  is necessary for the correct assignment of  $\text{CH}_4$  production mechanisms and oxidation based on  $\delta\text{D}$  and  $\delta^{13}\text{C}$  of  $\text{CH}_4$ . Samples collected for  $\delta\text{D}-\text{H}_2\text{O}$  were measured on an LGR DT-100 liquid water stable isotope analyser at Florida Agricultural and Mechanical University (Tallahassee, FL). Data analysis for these samples was performed with an MS Excel template from the IAEA Water Resources Programme (<http://www.iaea.org/water>).

Significant differences in  $\alpha_{\text{C}}$  and  $\delta\text{D}$  and  $\delta^{13}\text{C}$  of porewater  $\text{CH}_4$  between the *Sphagnum* and *Eriophorum* sites were determined with Student's *t*-test ( $\alpha_{\text{C}}$ ,  $\delta\text{D}-\text{CH}_4$ ,  $\delta^{13}\text{C}-\text{CH}_4$ ) and Hotelling's *t*-test (multivariate  $\delta\text{D}$  and  $\delta^{13}\text{C}$  of  $\text{CH}_4$ ) in R<sup>48</sup>. Statistical significance was determined at  $\alpha = 0.05$ .

**Peat sampling.** Peat samples were collected on 12 July 2011, 16 August 2011 and 16 October 2011 at three locations adjacent to the Palsa, *Sphagnum* and *Eriophorum* autochamber sites. For the *Sphagnum* and *Eriophorum* sites, samples were collected at the same depths and locations as those used for porewater sampling (Extended Data Table 1); sample depths for the Palsa site are detailed in ref. 6. Peat cores were collected with a push corer 11 cm in diameter (Palsa and *Sphagnum* sites) or a 10 cm  $\times$  10 cm Wardenaar corer (*Eriophorum* site). Cores were subsampled by depth and were subdivided in the field for microbial and chemical analysis, avoiding the outer 1 cm of the core. Samples for microbial analysis were placed in

cryotubes, saturated with about 3 volumes of LifeGuard solution (MoBio Laboratories) and stored at  $-80^\circ\text{C}$  until processing. Samples for chemical analysis were placed in plastic bags and frozen until processing.

**Peat chemical analysis.** For peat %C, %N, C:N ratio and  $\delta^{13}\text{C}$  measurements, 5–10 g of peat was dried at 60 °C until completely dry (3–10 days) and ground to a fine powder. Subsamples of ground peat (80–100  $\mu\text{g}$  for %C and  $\delta^{13}\text{C}$  analysis, and 5–6 mg for %N analysis) were wrapped in tin capsules and analysed by combustion to  $\text{CO}_2$  and  $\text{N}_2$  at 1,020 °C in an automated CHN elemental analyser coupled with a ThermoFinnigan Delta XP isotope ratio mass spectrometer at the National High Magnetic Field Laboratory. Samples were run in non-dilution mode for carbon analysis and dilution mode ( $\times 10$ ) for nitrogen analysis. C:N was calculated as  $(\%C)/(\%N)$  (by weight) for corresponding pairs of subsamples.

**Small-subunit ribosomal RNA gene amplicon analysis.** Sampling and extraction was performed as described previously<sup>6</sup>. Several additional samples were analysed for this paper; multiplex identifiers for those runs not reported in ref. 6 are provided in Extended Data Table 7. Small-subunit rRNA gene sequences were processed with APP 3.0.3 (<https://github.com/ECogenomics/APP>). Homopolymer errors were corrected with Acacia<sup>51</sup> and the resulting reads were processed by using the CD-HIT-OTU 0.0.2 pipeline with minor adjustments<sup>52</sup>. All reads were trimmed to 250 base pairs, and reads of less than 250 base pairs were discarded. Sequences were clustered at 97% identity and each cluster was assigned a taxonomy using BLASTN 2.2.22 (ref. 53) through the QIIME script `assign_taxonomy.py`<sup>54</sup> against the GreenGenes October 2012 database clustered at 99% identity (Supplementary Table 1). The taxonomy of each methanogenic cluster was confirmed by using parsimony insertion in ARB<sup>55</sup>. Amplicon sequence clusters were identified as potential hydrogenotrophic or acetoclastic methanogens based on taxonomic relationship to known methanogenic lineages (Extended Data Table 2)<sup>23,24,56</sup>. Within the order *Methanosarcinales*, lineages most closely related to *Methanoseta* were classified as obligate acetoclasts, whereas those most closely related to *Methanosarcina* were considered facultative acetoclasts, having the potential for both acetoclastic or hydrogenotrophic production<sup>23</sup>.

**Regression analysis.** A stepwise regression approach with Akaike's information criterion (AIC) as the model selection criterion was used to identify a subset of microbial and environmental predictor variables that best explained  $\text{CH}_4$  metabolism patterns quantified as porewater  $\alpha_{\text{C}}$  (Extended Data Table 5). Model selection was performed with the stepAIC package in R, and the relative importance of the predictor variables in the selected model was then calculated with the `relaimpo` R package<sup>48</sup>. Variables included in the model selection process included the relative abundances of the six most abundant methanogen operational taxonomic units (comprising more than 93% of the total methanogen sequences; see Extended Data Table 2) plus soil temperature, water table depth, pH, porewater  $\text{CH}_4$  and DIC concentration, and peat C:N, %C, %N and  $\delta^{13}\text{C}$  (Extended Data Table 1). Strong correlation between pH and both water-table depth and peat  $\delta^{13}\text{C}$  as well as peat %N and both %C and C:N meant that pH and %N were excluded from the regression analysis. Removal of non-significant predictor variables (DIC and relative abundance of an unidentified *Methanobacterium* spp. (otu-3636; Extended Data Table 2)) had a minimal effect on the model AIC value (less than 1); this simplified version was therefore selected as the optimal model (model 2 in Extended Data Table 5). Stepwise regression was also performed with  $\delta^{13}\text{C}-\text{CH}_4$  as the dependent variable. This analysis resulted in a similar model outcome, but with a lower  $R^2$  (model 1 in Extended Data Table 8). Stepwise regression analysis with environmental predictor variables and the relative abundance of the influential methanogen '*M. stordalenmirens*' (otu-10747) as the dependent variable showed that patterns in this methanogen's abundance were influenced by environmental conditions, particularly water table depth and peat chemistry (model 2 in Extended Data Table 8). However, these environmental variables alone cannot fully replace microbial data when modelling  $\alpha_{\text{C}}$ . Stepwise regression analysis using only environmental variables to predict  $\alpha_{\text{C}}$  yielded a model with a lower AIC and  $R^2$  (model 3 in Extended Data Table 8). It is the combination of methanogen and environmental variables that yields a model that explains the most variability in  $\alpha_{\text{C}}$  (Extended Data Table 5).

**Box model of atmospheric methane.** The model used here was a one-box model simplified from the two-box model of ref. 57 (and also used in the methane inversion study<sup>28</sup>):

$$\frac{dM}{dt} = F_{\text{CH}_4} - \lambda M \quad (3)$$

$$\frac{d(RM)}{dt} = R_{\text{CH}_4}F_{\text{CH}_4} - \alpha_{\text{OH}}\lambda(RM)$$

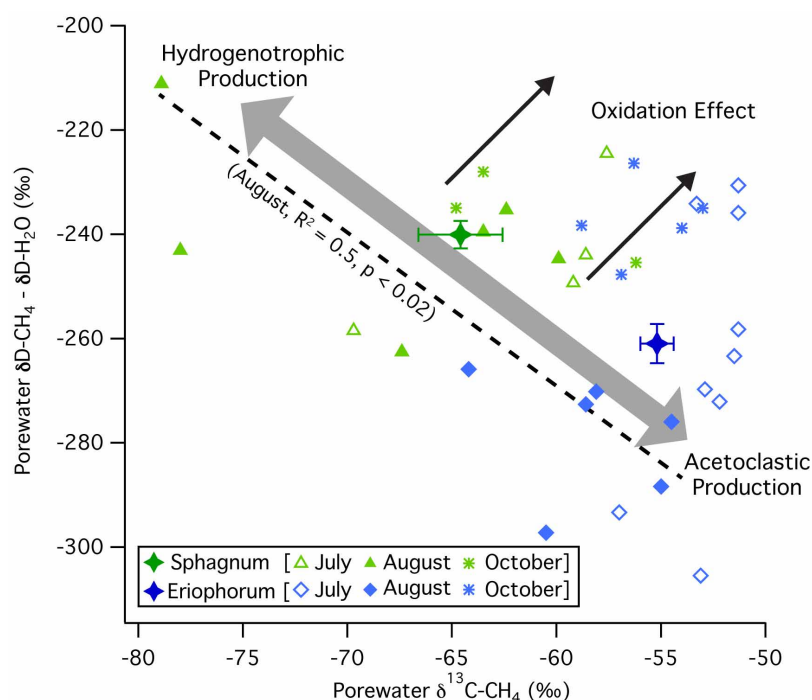
where  $M$  is the mixing ratio (in p.p.b.v.) of  $\text{CH}_4$  in the atmosphere,  $F_{\text{CH}_4}$  is the source flux of  $\text{CH}_4$  to the atmosphere,  $\lambda$  is the atmospheric removal rate ( $1/9\text{ yr}^{-1}$ , assumed for this illustration to be fixed), the  $R$  terms are the ratio of  $^{13}\text{CH}_4$  to  $^{12}\text{CH}_4$ , as defined for equation (1), and  $\alpha_{\text{OH}}$  is the isotopic fractionation (0.994, or

about  $-6\text{‰}$  for the atmospheric oxidation of  $\text{CH}_4$  by OH (ref. 28). Baseline flux to the atmosphere ( $F_{\text{CH}_4}$ ) was set to 559 Tg  $\text{CH}_4$ , the 1980 value<sup>28</sup>. The isotopic composition of  $\text{CH}_4$  inputs to the atmosphere ( $R_{\text{CH}_4}$ ) was set to the equivalent of  $-53\text{‰}$  to allow steady-state modern atmospheric  $\text{CH}_4$  to have the observed value of about  $-47\text{‰}$ .

We implemented this model numerically in the R software package<sup>48</sup>, simulating the effect on the atmosphere of  $\text{CH}_4$  emission due to permafrost thaw and partial decomposition of the 1,700 Pg C stock of permafrost C expected over the next 300 years, as summarized in refs 1, 2. High and low permafrost carbon release scenarios for both the high climate change scenario (Intergovernmental Panel on Climate Change (IPCC) scenario RCP8.5, leading to the release of 120–195 Pg C) and the low climate change scenario (IPCC scenario RCP2.6, approximated as one-third of the C release of the high scenario) (Extended Data Fig. 2a) generated  $\text{CH}_4$  emissions (Fig. 3a) (based on 2.3% of released permafrost carbon emerging as  $\text{CH}_4$  (ref. 2)) and corresponding impacts on the atmospheric concentrations of  $\text{CH}_4$  (Extended Data Fig. 2b). We simulated the impacts of these emissions on the isotopic composition of atmospheric  $\text{CH}_4$  by assuming that the  $\delta^{13}\text{C}$  of  $\text{CH}_4$  emitted was in the range of what we report here for Stordalen mire, from very light ( $-80\text{‰}$ , like that measured at the *Sphagnum* site) to only moderately light ( $-65\text{‰}$ , like that measured at the *Eriophorum* site), giving a range of isotopic perturbations to atmospheric  $\text{CH}_4$  under high climate change (Extended Data Fig. 2c) and under low climate change (Extended Data Fig. 2d). In all scenarios, the induced change in atmospheric  $\delta^{13}\text{C}$  is significantly larger than the atmospheric detection limit of 0.1‰ (reported in ref. 28 and shown as a dotted horizontal line in Extended Data Fig. 2c, d).

For the analysis shown in Fig. 3 we focused on a mid-range value of permafrost C release (high climate change scenario with low C release, 120 Pg total C by 2100), corresponding to emissions of 2.8 Pg C as  $\text{CH}_4$  by 2100 (the dashed black and red line in Fig. 3a). (By comparison, the IPCC estimates that up to 5 Pg C may be released as  $\text{CH}_4$  by 2100 (ref. 3).) We explored the misattribution of C release that would occur, by (mistakenly) assuming that the isotopic composition of emitted  $\text{CH}_4$  was in the range of assumptions used in previous atmospheric inversions, from  $-60\text{‰}$  to  $-65\text{‰}$  (ref. 28), instead of the range measured at Stordalen mire ( $-65\text{‰}$  to  $-80\text{‰}$ ). We estimated the magnitude of misattribution (or error flux; Fig. 3c) by simulating the amount of additional carbon that would need to be released (at nominally assumed isotopic composition values of  $-60$  or  $-65$ ) to have the same effect on atmospheric composition as the carbon released under scenarios with isotopic compositions like those observed in the field.

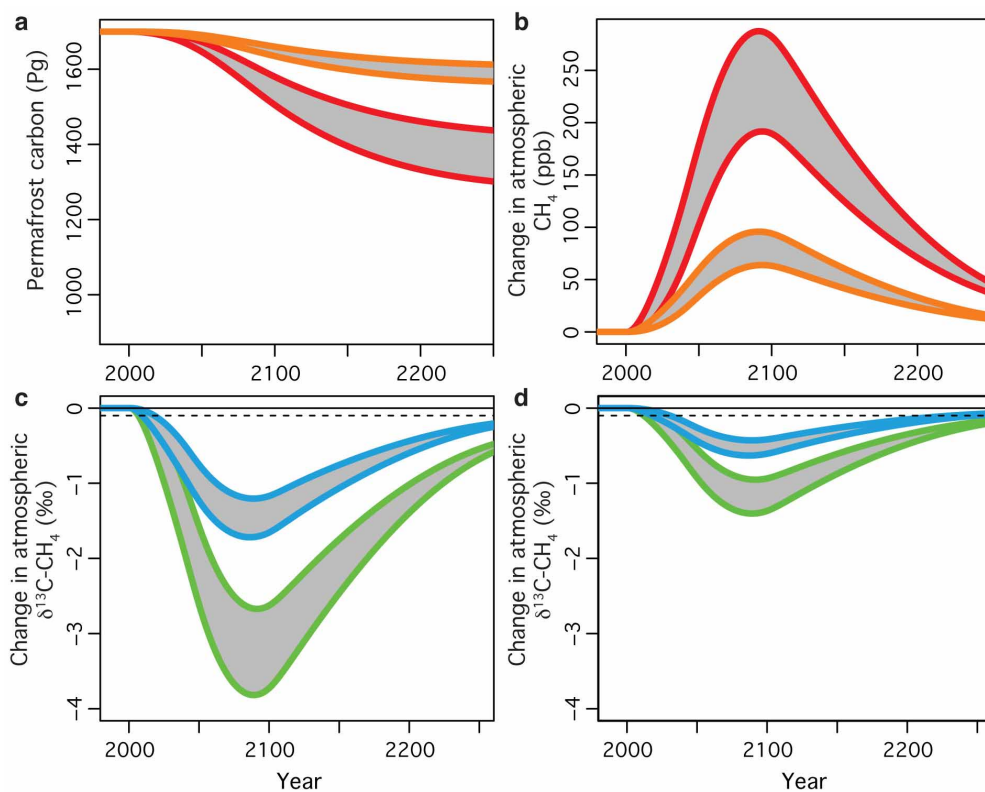
31. Payette, S. Accelerated thawing of subarctic peatland permafrost over the last 50 years. *Geophys. Res. Lett.* **31**, 1–4 (2004).
32. O'Donnell, J. a. et al. The effects of permafrost thaw on soil hydrologic, thermal, and carbon dynamics in an Alaskan peatland. *Ecosystems* **15**, 213–229 (2012).
33. Vitt, D. H., Halsey, L. A. & Zoltai, S. C. The changing landscape of Canada's western boreal forest: the current dynamics of permafrost. *Can. J. For. Res.* **30**, 283–287 (2000).
34. Quinton, W. L., Hayashi, M. & Chasmer, L. E. Permafrost-thaw-induced land-cover change in the Canadian subarctic: implications for water resources. *Hydrol. Processes* **25**, 152–158 (2011).
35. Zoltai, S. C. Cyclic development of permafrost in the peatlands of Northwestern Alberta, Canada. *Arct. Alp. Res.* **25**, 240–246 (1993).
36. Camill, P. & Clark, J. S. Climate change disequilibrium of boreal permafrost peatlands caused by local processes. *Am. Nat.* **151**, 207–222 (1998).
37. Dyke, L. D. & Sladen, W. E. Permafrost and peatland evolution in the Northern Hudson Bay Lowland, Manitoba. *Arctic* **63**, 429–441 (2010).
38. Whiticar, M. J. & Faber, E. Methane oxidation in sediment and water column environments—isotopic evidence. *Org. Geochem.* **10**, 759–768 (1986).
39. Conrad, R. Quantification of methanogenic pathways using stable carbon isotopic signatures: a review and a proposal. *Org. Geochem.* **36**, 739–752 (2005).
40. Bäckstrand, K., Crill, P. M., Mastepanov, M., Christensen, T. R. & Bastviken, D. Non-methane volatile organic compound flux from a subarctic mire in Northern Sweden. *Tellus B Chem. Phys. Meteorol.* **60**, 226–237 (2008).
41. Bubier, J. L., Crill, P. M., Mosedale, A., Frolking, S. & Linder, E. Peatland responses to varying interannual moisture conditions as measured by automatic  $\text{CO}_2$  chambers. *Glob. Biogeochem. Cycles* **17**, 1066, <http://dx.doi.org/10.1029/2002GB001946> (2003).
42. Santoni, G. W. et al. Mass fluxes and isofluxes of methane ( $\text{CH}_4$ ) at a New Hampshire fen measured by a continuous wave quantum cascade laser spectrometer. *J. Geophys. Res.* **117**, D10301 (2012).
43. Werle, P., Mücke, R. & Slemr, F. The limits of signal averaging in atmospheric trace-gas monitoring by tunable diode-laser absorption spectroscopy (TDLAS). *Appl. Phys. B* **139**, 131–139 (1993).
44. Bäckstrand, K., Crill, P. M., Mastepanov, M., Christensen, T. R. & Bastviken, D. Total hydrocarbon flux dynamics at a subarctic mire in northern Sweden. *J. Geophys. Res.* **113**, G03026, <http://dx.doi.org/10.1029/2008JG000703> (2008).
45. Pataki, D. E. The application and interpretation of Keeling plots in terrestrial carbon cycle research. *Glob. Biogeochem. Cycles* **17**, 1022, <http://dx.doi.org/10.1029/2001GB001850> (2003).
46. Keeling, C. D. The concentration and isotopic abundances of atmospheric carbon dioxide in rural areas. *Geochim. Cosmochim. Acta* **13**, 322–334 (1958).
47. Keeling, C. D. The concentration and isotopic abundances of carbon dioxide in rural and marine air. *Geochim. Cosmochim. Acta* **24**, 277–298 (1960).
48. R Development Core Team. *A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2012).
49. Corbett, J. E. et al. Partitioning pathways of  $\text{CO}_2$  production in peatlands with stable carbon isotopes. *Biogeochemistry* **114**, 327–340 (2013).
50. Chanton, J. P., Fields, D. & Hines, M. E. Controls on the hydrogen isotopic composition of biogenic methane from high-latitude terrestrial wetlands. *J. Geophys. Res.* **111**, 1–9 (2006).
51. Bragg, L., Stone, G., Imelfort, M., Hugenholtz, P. & Tyson, G. W. Fast, accurate error-correction of amplicon pyrosequences. *Nature Methods* **9**, 425–426 (2012).
52. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of proteins or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).
53. Altschul, H. J. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
54. Caporaso, J. G. et al. QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7**, 335–336 (2010).
55. Ludwig, W. et al. ARB: a software environment for sequence data. *Nucleic Acids Res.* **32**, 1363–1371 (2004).
56. Conrad, R. The global methane cycle: recent advances in understanding the microbial processes involved. *Environ. Microbiol. Rep.* **1**, 285–292 (2009).
57. Tans, P. P. A note on isotopic ratios and the global atmospheric methane budget. *Glob. Biogeochem. Cycles* **11**, 77–81 (1997).



**Extended Data Figure 1 | Expected and observed relationships between the  $\delta\text{D}$  and  $\delta^{13}\text{C}$  content of porewater  $\text{CH}_4$ .** The thick grey arrow shows the expected pattern in H and C isotopes of  $\text{CH}_4$  when variations are caused by shifts between acetoclastic (lower right) and hydrogenotrophic (upper left) production. The thin black arrows pointing to the upper right indicate the expected pattern in H and C isotopes of  $\text{CH}_4$  when variations are caused by changes in  $\text{CH}_4$  oxidation<sup>19</sup>. The points are observed isotopic compositions of samples collected between July and October 2011 at the partly thawed *Sphagnum* and fully thawed *Eriophorum* sites; site averages are shown with error bars (error bars represent s.e.m.;  $n = 13$  (*Sphagnum*) and 20 (*Eriophorum*)). Although the scatter allows for some variation in both

production and oxidation, the average *Eriophorum* porewater  $\text{CH}_4$  had significantly more  $^{13}\text{C}$  and less D relative to *Sphagnum* porewater (Hotelling's  $T^2$  test,  $P = 0.0001$ ,  $n = 33$ ), indicating that the overall inter-site isotopic differences were due mostly to differences in the  $\text{CH}_4$  production pathway rather than to differences in  $\text{CH}_4$  oxidation. Additionally, in August there was a significant negative relationship between  $\delta^{13}\text{C}-\text{CH}_4$  and  $\delta\text{D}-\text{CH}_4$  of porewater samples collected across sites (dashed line, linear regression,  $R^2 = 0.5$ ,  $P < 0.02$ ,  $n = 12$ ). Note that on the vertical axis  $\delta\text{D}-\text{H}_2\text{O}$  has been subtracted from  $\delta\text{D}-\text{CH}_4$  to correct for the effect of  $\delta\text{D}$  exchange between  $\text{H}_2\text{O}$  and  $\text{CH}_4$  (refs 20, 38, 50).





**Extended Data Figure 2 | Simulations, using high and low temperature and C release scenarios, of the effect of CH<sub>4</sub> release from thawing permafrost on atmospheric δ<sup>13</sup>C-CH<sub>4</sub>.** **a**, Scenarios of permafrost C release due to thaw (red bounding lines, high temperature; orange bounding lines, low temperature; the range in each case is defined by high and low C release scenarios). **b**, Impact on atmospheric methane mixing ratios (assuming that 2.3% of released C is

emitted as methane). **c**, Impact of the high climate change scenario on atmospheric methane isotopes, assuming *Eriophorum*-like emissions (blue bounding lines, δ<sup>13</sup>C ≈ -65‰), or assuming *Sphagnum*-like emissions (green bounding lines, δ<sup>13</sup>C ≈ -80‰). **d**, As in **c**, except for the low climate change scenario. In **c** and **d**, dotted horizontal lines indicate the detection limit for CH<sub>4</sub> isotopes<sup>28</sup>.

Extended Data Table 1 | Summary of porewater chemistry, average (s.e.m.),  $n = 3$ 

Sample	Depth (cm)	pH	mM CO <sub>2</sub>	mM CH <sub>4</sub>	$\delta^{13}\text{CO}_2$ ‰	$\delta^{13}\text{CH}_4$ ‰	$\alpha_c$
July, 2011							
<i>Sphagnum</i> - M	13	4.1 (0.06)	3.02 (0.78)	0.09 (0.04)	-15.7 (1.6)	-62.2 (3.8)	1.050 (0.005)
<i>Sphagnum</i> - D	19	4.2 (0.09)	3.50 (0.57)	0.15 (0.05)	-14.1 (0.6)	-62.2 (4.5)	1.051 (0.005)
<i>Eriophorum</i> - S	3	5.8 (0.09)	2.29 (0.92)	0.18 (0.12)	-14.1 (1.1)	-52.1 (0.5)	1.040 (0.001)
<i>Eriophorum</i> -M	7	5.6 (0.06)	3.06 (0.77)	0.28 (0.07)	-12.9 (1.0)	-52.6 (0.6)	1.042 (0.001)
<i>Eriophorum</i> - D	24	5.6 (0.03)	3.56 (0.80)	0.36 (0.07)	-11.6 (1.7)	-53.3 (1.9)	1.044 (0.004)
August, 2011							
<i>Sphagnum</i> - M	21	4.2 (0.10)	4.89 (0.37)	0.23 (0.04)	-12.0 (1.5)	-66.7 (5.7)	1.059 (0.008)
<i>Sphagnum</i> - D	26	4.1 (0.13)	4.80 (0.48)	0.23 (0.04)	-10.7 (1.6)	-69.9 (4.6)	1.064 (0.007)
<i>Eriophorum</i> - S	3	5.7 (0.19)	1.62 (0.28)	0.06 (0.04)	-13.5 (0.5)	-60.0 (2.6)	1.049 (0.003)
<i>Eriophorum</i> -M	7	5.7 (0.10)	1.93 (0.25)	0.10 (0.02)	-13.9 (0.4)	-56.6 (2.1)	1.045 (0.002)
<i>Eriophorum</i> - D	26	5.6 (0.15)	3.58 (0.62)	0.31 (0.11)	-11.1 (2.4)	-55.9 (1.1)	1.047 (0.001)
October, 2011							
<i>Sphagnum</i> - M	10	4.3 (0.06)	1.24 (0.42)	0.03 (0.02)	-16.4 (1.6)	-59.2 (6.5)	1.046 (0.006)
<i>Sphagnum</i> - D	15	4.5 (0.10)	3.21 (0.90)	0.10 (0.04)	-13.8 (2.4)	-61.5 (2.7)	1.051 (0.0004)
<i>Eriophorum</i> - S	3	5.9 (0.15)	2.15 (1.43)	0.19 (0.13)	-14.1 (1.0)	-56.4 (2.4)	1.045 (0.001)
<i>Eriophorum</i> -M	7	5.9 (0.15)	2.71 (1.25)	0.29 (0.14)	-13.7 (1.6)	-57.8 (3.1)	1.047 (0.002)
<i>Eriophorum</i> - D	26	5.7 (0.12)	3.84 (1.64)	0.53 (0.27)	-11.3 (3.1)	-58.1 (2.2)	1.050 (0.001)

**Extended Data Table 2 | Relative abundance, taxonomic classification and predicted methanogenic pathway of the dominant methanogen operational taxonomic units (OTUs)**

Sample	Candidatus <i>Methanoflorens</i> (otu-10747) Hydrogenotrophic	<i>Methanobacterium</i> (otu-3636) Hydrogenotrophic	Candidatus <i>Methanoregula</i> (otu-20819) Hydrogenotrophic	<i>Methanosarcina</i> (otu-7308) Acetoclastic (facultative)	<i>Methanosaeta</i> (otu-10220) Acetoclastic (obligate)	<i>Methanosaeta</i> (otu-15150) Acetoclastic (obligate)
<b>July, 2011</b>						
Palsa – S	0.0	0.0	0.0	0.0	0.0	0.0
Palsa – M	0.0	0.0	0.0	0.0	0.0	0.0
Palsa – D	0.0	0.4	0.0	0.0	0.0	0.0
<i>Sphagnum</i> – S	0.3	0.4	0.0	0.1	0.0	0.0
<i>Sphagnum</i> – M	4.0	12.9	0.0	3.4	0.0	0.0
<i>Sphagnum</i> – D	16.4	5.8	0.0	3.3	0.0	0.0
<i>Eriophorum</i> – S	1.0	2.7	5.8	0.7	4.5	1.8
<i>Eriophorum</i> – M	5.3	3.7	4.0	2.2	5.0	2.7
<i>Eriophorum</i> – D	8.3	1.6	1.9	0.6	4.2	1.2
<b>August, 2011</b>						
Palsa – S	0.0	0.0	0.0	0.0	0.0	0.0
Palsa – M	0.0	0.0	0.0	0.0	0.0	0.0
Palsa – D	0.0	0.0	0.0	0.0	0.0	0.0
<i>Sphagnum</i> – S	0.1	0.4	0.0	0.2	0.0	0.0
<i>Sphagnum</i> – M	11.6	4.0	0.0	1.9	0.0	0.0
<i>Sphagnum</i> – D	32.1	3.1	0.0	1.4	0.0	0.0
<i>Eriophorum</i> – S	0.6	2.1	3.6	0.4	3.3	1.0
<i>Eriophorum</i> – M	6.3	6.1	5.1	2.6	9.0	3.9
<i>Eriophorum</i> – D	6.5	0.3	3.4	1.2	1.7	0.6
<b>October, 2011</b>						
Palsa – S	0.0	0.0	0.0	0.0	0.0	0.0
Palsa – M	0.1	1.1	0.0	0.1	0.0	0.0
Palsa – D	0.1	0.7	0.0	0.0	0.0	0.0
<i>Sphagnum</i> – S	0.0	0.1	0.0	0.0	0.0	0.0
<i>Sphagnum</i> – M	0.0	3.4	0.0	1.1	0.0	0.0
<i>Sphagnum</i> – D	0.6	8.4	0.0	1.2	0.0	0.0
<i>Eriophorum</i> – S	2.5	1.7	1.7	0.6	1.4	0.6
<i>Eriophorum</i> – M	2.1	1.9	1.0	0.8	2.5	2.2
<i>Eriophorum</i> – D	6.0	1.1	3.7	0.1	5.1	5.8



**Extended Data Table 3 | Relative abundance of methanogen functional groups within the Archaea**

Site	Hydrogenotrophic	Acetoclastic (facultative)	Acetoclastic (obligate)	Other Archaea
<b>July, 2011</b>				
Palsa	35.9	2.9	0.0	61.2
<i>Sphagnum</i> (aerobic)*	83.1	15.5	0.0	1.4
<i>Sphagnum</i> (anaerobic)†	82.1	14.2	0.0	3.8
<i>Eriophorum</i>	39.5	4.2	21.4	34.9
<b>August, 2011</b>				
Palsa	0.0	8.7	0.0	91.3
<i>Sphagnum</i> (aerobic)*	68.2	30.7	0.0	1.1
<i>Sphagnum</i> (anaerobic)†	91.2	6.1	0.0	2.8
<i>Eriophorum</i>	39.5	5.1	21.9	33.5
<b>October, 2011</b>				
Palsa	56.5	2.6	0.4	40.5
<i>Sphagnum</i> (aerobic)*	65.7	24.0	0.7	9.6
<i>Sphagnum</i> (anaerobic)†	15.6	2.8	2.6	79.0
<i>Eriophorum</i>	35.8	2.4	27.6	34.2

\* Above the water table.

† Below the water table.

**Extended Data Table 4 | Results of linear regression analysis for predicting  $\alpha_C$  from the relative abundances of methanogenic pathways, dominant methanogenic lineages and environmental variables ( $n = 41$ )**

Variable	R <sup>2</sup>	F-statistic	p-value
' <i>M. stordalenmirensis</i> '	0.58	54.09	<0.001
otu-3636*	0.00	0.01	0.926
otu-10220*	0.12	5.36	0.026
otu-20819 *	0.15	6.82	0.013
otu-15150 *	0.06	2.27	0.140
otu-7308 *	0.01	0.32	0.576
Hydrogenotrophic	0.44	30.63	<0.001
Acetoclastic (obligate)	0.12	5.23	0.028
Water table depth	0.44	31.1	<0.001
pH	0.19	8.97	0.005
Porewater CH <sub>4</sub> (mM)	0.00	0.07	0.796
Porewater DIC (mM)	0.25	13.33	0.001
Peat C:N	0.00	0.17	0.682
Peat %C	0.02	0.75	0.393
Peat %N	0.00	0.14	0.709
Peat $\delta^{13}\text{C}$	0.13	5.99	0.019

\* See Extended Data Table 2 for taxonomic details.

**Extended Data Table 5 | Results of stepwise multiple regression analysis for predicting  $\alpha_C$  from relative abundances of methanogenic lineages and environmental variables**

Variable	Coefficient	Std Error	t value	p value	Cumulative AIC
<b>Model 1 - stepwise regression, direction = both</b> ( $R^2 = 0.81$ , $F = 23.71$ on 6 and 34 df, $p < 0.001$ )					
Water table depth	-0.0004	0.0001	-5.398	<0.001	-422.33
' <i>M. stordalenmirens</i> '	0.0271	0.0084	3.221	0.002	-436.79
C:N	-0.0002	0.0001	-2.872	0.007	-438.80
Peat $\delta^{13}\text{C}$	0.0014	0.0006	2.516	0.017	-440.71
DIC (mM)	0.0007	0.0005	1.396	0.171	-445.42
otu-3636*	-0.0271	0.0161	-1.345	0.188	-445.58
Intercept	1.089	0.0167	65.193	<0.001	-445.71
<b>Model 2 – significant predictor variables from model 1</b> ( $R^2 = 0.79$ , $F = 33.71$ on 4 and 36 df, $p < 0.001$ )					
Water table depth	-0.0004	0.0001	-5.202	<0.001	-425.11
' <i>M. stordalenmirens</i> '	0.0351	0.0072	4.867	<0.001	-427.36
C:N	-0.0002	0.0001	-2.613	0.013	-440.97
Peat $\delta^{13}\text{C}$	0.0014	0.0006	2.470	0.018	-441.67
Intercept	1.089	0.0164	66.583	<0.001	-446.09

\* See Extended Data Table 2 for taxonomic details.



**Extended Data Table 6 | Estimate of the relative contribution of hydrogenotrophic production to annual CH<sub>4</sub> emission at Stordalen mire**

Habitat	Area (ha)*	Annual Flux (g CH <sub>4</sub> m <sup>-2</sup> ) <sup>†</sup>	Annual Emission (kg CH <sub>4</sub> ) <sup>*,†</sup>	Estimated Emission from Hydrogenotrophy (kg CH <sub>4</sub> yr <sup>-1</sup> ) <sup>‡</sup>
<i>Sphagnum</i>	6.2	6.2	288.3	247.9 <sup>§</sup> - 282.5 <sup>  </sup>
<i>Eriophorum</i>	2.0	36.0	540.6	172.8 <sup>¶</sup> - 335.2 <sup>#</sup>
<b>Total</b>			<b>828.9</b>	<b>420.7(51%) - 617.7 (75%)</b>

\* Based on ref. 4; the *Sphagnum* site in this study is representative of the semi-wet and wet vegetation classes.

† Annual total hydrocarbon emissions from ref. 16 corrected for non-methane volatile organic compound (NMVOC) flux using the reported proportions (25% NMVOC for the *Eriophorum* site; 15% for the *Sphagnum* site). The magnitude of growing season CH<sub>4</sub> emissions measured in this study is comparable to the growing season CH<sub>4</sub> flux used in the estimate of annual flux in ref. 16.

‡ Two approaches: isotopic, using mixing of acetoclastic (−60‰) and hydrogenotrophic (−80‰) sources to yield mean emitted δ<sup>13</sup>C-CH<sub>4</sub>, and molecular, using the proportion of the methanogen community identified as hydrogenotrophic.

§ Molecular approach: on average 86% of methanogen community in the anoxic CH<sub>4</sub>-producing peat was identified as hydrogenotrophic; all of the acetoclasts were facultative so this is probably an underestimate of potential hydrogenotrophic production.

|| Isotopic approach: −79.6‰ = −80‰ × **0.98** + −60‰ × 0.02 (the bold number indicates the proportion of CH<sub>4</sub> produced by hydrogenotrophy that would produce the measured δ<sup>13</sup>C-CH<sub>4</sub>).

¶ Isotopic approach: −66.3‰ = −80‰ × **0.32** + −60‰ × 0.68 (the bold number indicates the proportion of CH<sub>4</sub> produced by hydrogenotrophy that would produce the measured δ<sup>13</sup>C-CH<sub>4</sub>).

#Molecular approach: on average 62% of the methanogen community was identified as hydrogenotrophic.

**Extended Data Table 7 | Small-subunit rRNA gene amplicon multiplex identifiers (MIDs) used for each sample**

Sample name	Run #	Multiplex identifier (MID)
20110712_E_3_M	6	CGAGC
20110712_S_1_M	6	CGCAT
20110712_S_3_M	6	CGTAC
20110712_P_1_S	6	CGTGT
20110712_P_2_S	6	CTAGT
20110712_P_3_S	6	CTGAC
20110816_S_2_S	6	TACGC
20110816_S_1_D	6	TATGT
20110816_P_1_M	6	TCAGT
20111016_P_1_S	6	TCGAT

\* Sample names are composed of the date of sampling, followed by P, S or E for Palsa, *Sphagnum* or *Eriophorum* sites, respectively; the number indicates the core within the site, and S, M or D indicates surface, middle or deep sampling within the core, respectively.

† Samples were multiplexed in six separate runs, each time with samples not related to this study. The multiplex identifiers of the first five runs are given in ref. 6.

**Extended Data Table 8 | Results of stepwise multiple regression analysis for predicting  $\delta^{13}\text{C}\text{-CH}_4$  from relative abundances of methanogenic lineages and environmental variables (model 1), the relative abundance of '*M. stordalenmirens*' from environmental variables (model 2), and  $\alpha_{\text{C}}$  from environmental variables (model 3)**

Variable	Coefficient	Std Error	t value	p value	Cumulative AIC
<b>Model 1</b> - stepwise regression, dependent variable = $\delta^{13}\text{C}\text{-CH}_4$ , direction = both $(R^2 = 0.75, F = 21.25 \text{ on } 5 \text{ and } 35 \text{ df, } p < 0.001)$					
Water table depth	0.299	0.07	4.512	<0.001	130.95
' <i>M. stordalenmirens</i> '	-23.25	6.79	-3.426	0.002	124.01
Peat $\delta^{13}\text{C}$	-1.51	0.54	-2.779	0.009	120.33
$\text{CH}_4$ (mM)	10.60	4.12	2.576	0.014	119.28
C:N	0.12	0.05	2.149	0.039	117.24
Intercept	-102.14	15.23	-6.705	<0.001	114.16
<b>Model 2</b> - stepwise regression, dependent variable = ' <i>M. stordalenmirens</i> ', direction = both $(R^2 = 0.53, F = 7.77 \text{ on } 5 \text{ and } 35 \text{ df, } p < 0.001)$					
Water table depth	-0.0053	0.0015	-3.634	<0.001	-188.03
C:N	-0.0035	0.0010	-3.495	0.001	-188.88
DIC (mM)	0.0214	0.0106	2.025	0.050	-196.61
% C	0.0033	0.0018	1.799	0.081	-197.53
Soil temperature	0.0059	0.0040	1.483	0.147	-198.66
Intercept	-0.0558	0.0805	-0.692	0.493	-199.15
<b>Model 3</b> - stepwise regression, dependent variable = $\alpha_{\text{C}}$ , direction = both $(R^2 = 0.71, F = 21.71 \text{ on } 4 \text{ and } 36 \text{ df, } p < 0.001)$					
Water table depth	-0.0005	0.0001	-6.465	<0.001	-402.97
C:N	-0.0003	0.0001	-4.514	<0.001	-416.18
DIC (mM)	0.0015	0.0006	2.629	0.013	-427.36
Peat $\delta^{13}\text{C}$	0.0017	0.0007	2.574	0.014	-427.63
Intercept	1.0990	0.0192	57.396	<0.001	-432.56



# Limited impact on decadal-scale climate change from increased use of natural gas

Haewon McJeon<sup>1</sup>, Jae Edmonds<sup>1</sup>, Nico Bauer<sup>2</sup>, Leon Clarke<sup>1</sup>, Brian Fisher<sup>3</sup>, Brian P. Flannery<sup>4</sup>, Jérôme Hilaire<sup>2</sup>, Volker Krey<sup>5</sup>, Giacomo Marangoni<sup>6</sup>, Raymond Mi<sup>3</sup>, Keywan Riahi<sup>5</sup>, Holger Rogner<sup>5</sup> & Massimo Tavoni<sup>6</sup>

The most important energy development of the past decade has been the wide deployment of hydraulic fracturing technologies that enable the production of previously uneconomic shale gas resources in North America<sup>1</sup>. If these advanced gas production technologies were to be deployed globally, the energy market could see a large influx of economically competitive unconventional gas resources<sup>2</sup>. The climate implications of such abundant natural gas have been hotly debated. Some researchers have observed that abundant natural gas substituting for coal could reduce carbon dioxide (CO<sub>2</sub>) emissions<sup>3–6</sup>. Others have reported that the non-CO<sub>2</sub> greenhouse gas emissions associated with shale gas production make its lifecycle emissions higher than those of coal<sup>7,8</sup>. Assessment of the full impact of abundant gas on climate change requires an integrated approach to the global energy–economy–climate systems, but the literature has been limited in either its geographic scope<sup>9,10</sup> or its coverage of greenhouse gases<sup>2</sup>. Here we show that market-driven increases in global supplies of unconventional natural gas do not discernibly reduce the trajectory of greenhouse gas emissions or climate forcing. Our results, based on simulations from five state-of-the-art integrated assessment models<sup>11</sup> of energy–economy–climate systems independently forced by an abundant gas scenario, project large additional natural gas consumption of up to +170 per cent by 2050. The impact on CO<sub>2</sub> emissions, however, is found to be much smaller (from –2 per cent to +11 per cent), and a majority of the models reported a small increase in climate forcing (from –0.3 per cent to +7 per cent) associated with the increased use of abundant gas. Our results show that although market penetration of globally abundant gas may substantially change the future energy system, it is not necessarily an effective substitute for climate change mitigation policy<sup>9,10</sup>.

Five research teams projected the evolution of the future global energy system up to 2050 under two alternative assumptions about natural gas supply: ‘Conventional Gas’ and ‘Abundant Gas’ (Fig. 1 and Methods). Each natural gas supply curve was constructed based on the synthesis of natural gas supply and geographic distribution in the Global Energy Assessment (GEA) report<sup>12</sup>.

The Conventional Gas scenario assumes the maximum recoverable resources to be 11,000 exajoules (EJ) in 2010, a total consistent with conventional resources that have extraction costs below \$3 per gigajoule (GJ). (One EJ equals one quintillion (10<sup>18</sup>) joules and one GJ equals one billion (10<sup>9</sup>) joules.) This supply curve reflects an estimate of economically recoverable gas consistent with technology available before the shale gas revolution.

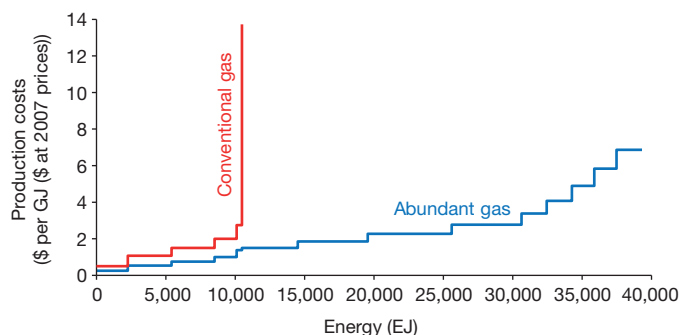
The Abundant Gas scenario is characterized by both the global abundance of natural gas resources and substantially reduced extraction costs. This scenario envisions that advanced natural gas extraction technologies become globally applicable beyond North America, allowing extraction of previously uneconomic unconventional resources. To represent this scenario, we assumed that technological change halves the extraction

cost in GEA between 2010 and 2050, allowing more than 30,000 EJ of cumulative natural gas to be produced at or below \$3 per GJ, with additional resources producible at higher prices. This rate of cost reduction is on the higher end compared to other studies<sup>2,9,10</sup>. This scenario is designed to provide a potential upper bound on global gas supply and should not be interpreted as the most likely case (see Methods for a broader range of supply assumptions). Furthermore, this rate of cost reduction is more aggressive than that of most low-carbon energy sources against which natural gas is competing (see Extended Data Table 1 for variance across the models).

For both scenarios, we did not simulate future climate policies beyond those already in effect. The two scenarios therefore explore the degree to which market penetration of abundant gas alone can mitigate greenhouse gas emissions (see Methods).

Five integrated assessment models (IAMs) are employed in this study: BAEGEM<sup>13</sup>, GCAM<sup>14</sup>, MESSAGE<sup>15</sup>, REMIND<sup>16</sup> and WITCH<sup>17</sup>. These IAMs belong to a class of models designed to assess the implications of changes in the global energy system on climate forcing. They have been used extensively to project emission scenarios for global and regional assessments. For example, GCAM and MESSAGE provided two of the four Representative Concentration Pathways (RCPs) used in the Intergovernmental Panel on Climate Change’s Fifth Assessment Report<sup>18,19</sup>.

The models integrate energy, economy, and climate systems to assess their interaction in a consistent framework. All models feature explicit representation of energy markets with price-responsive demand and supply for coal, oil, and gas, as well as for low-carbon energy sources<sup>20</sup>. The capability to simulate the effects of price changes on the scale and the composition of the future energy system is crucial for this study,



**Figure 1 | Global natural gas supply curves in 2050.** The current natural gas supply curves provided by Global Energy Assessment<sup>12</sup>. Future cost reduction assumptions are documented in the Methods. These supply costs are not the actual prices in the market place. The costs do not include taxes or royalties, nor do they include external environmental or social costs associated with gas production<sup>12</sup>. \$3 per GJ is equivalent to \$3.2 per mmBtu. (One mmBtu is one million British thermal units.) US dollars at 2007 constant prices.

<sup>1</sup>Pacific Northwest National Laboratory, JGCRI, 5825 University Research Court, Suite 3500, College Park, Maryland 20740, USA. <sup>2</sup>Potsdam Institute for Climate Impact Research, PO Box 60 12 03, D-14412 Potsdam, Germany. <sup>3</sup>BAEconomics, PO Box 5447, Kingston, Australian Capital Territory 2604, Australia. <sup>4</sup>Resources for the Future, 1616 P Street Northwest, Washington, DC 20036, USA.

<sup>5</sup>International Institute for Applied Systems Analysis, Schlossplatz 1, A-2361 Laxenburg, Austria. <sup>6</sup>Centro Euromediterraneo sui Cambiamenti Climatici and Politecnico di Milano, Via Lambruschini 4b, 20156 Milan, Italy.

because these effects determine the changes in emissions and corresponding changes in climate forcing.

The models are harmonized to share common natural gas supply curve assumptions, but otherwise differ widely in model architecture, geospatial resolution, socioeconomic assumptions, and technology projections (see Table 1 and Methods for a detailed description of the model differences and similarities). To the extent that a similar result is produced by this diverse set of models, we are more confident that the result is not simply an idiosyncratic artefact of an individual modelling method, but rather is reflective of more fundamental forces.

The models independently projected the future energy system for the two natural gas supply scenarios. All five models reported that the abundant gas supply leads to additional gas consumption, as well as additional gas-fired electricity consumption, compared to the Conventional Gas scenario. However, the speed of divergence and the size of the difference in gas consumption varied across models: from 11% in WITCH to 170% in REMIND in 2050 (Fig. 2a). The models agreed on the pattern of sector penetration. Power production showed the largest shift towards gas substituting for all other fuels, most prominently coal. Smaller shifts occurred in industry and buildings (Fig. 3a). The models also agreed that natural gas continues to have a minor role only in transportation.

Despite major changes to the global energy system and the substantial increase in natural gas consumption, the models agreed that additional supply of natural gas in the energy market does not discernibly reduce fossil fuel CO<sub>2</sub> emissions. Future CO<sub>2</sub> emissions are similar in magnitude with and without abundant gas, as the two emission trajectories continue to rise over time at similar rates (Fig. 2b). For GCAM, MESSAGE, and WITCH, the CO<sub>2</sub> emissions for both scenarios were within 2% of each other in 2050. The BAEGEM (11%) and REMIND (5%) models showed larger differences, but emissions increased—rather than decreased—under the Abundant Gas scenario.

The results demonstrate that abundant gas will not necessarily reduce CO<sub>2</sub> emissions. There are two forces at work: substitution and scale effect. First, additional natural gas consumption largely substitutes for coal, but not exclusively. All five models found that gas substitutes for all other primary fuels—such as nuclear and renewables—although coal loses the largest market share in all models (Fig. 3a). In 2050, abundant gas on average substitutes for 18% of coal and 17% of low-carbon energy (10% and 8% respectively for the 2010–2050 cumulative total). Hence, the effect of natural gas on CO<sub>2</sub> emissions is not based on the difference

between the emission factors of gas and of coal, but on the emission factor of gas relative to that of a broader basket of energy sources. The natural gas emission factor (56 kg of CO<sub>2</sub> per GJ) is about half of the coal emission factor (96 kg of CO<sub>2</sub> per GJ)<sup>21</sup>. However, it is not substantially lower than the average global CO<sub>2</sub> emissions per unit of energy: 68 kg of CO<sub>2</sub> per GJ (2050 model average). Consequently, even if natural gas were to substitute for the entire global energy supply, CO<sub>2</sub> emissions would decline by a maximum of 20%. Considering the model average of a 36% share for natural gas in the global energy system in 2050, the actual emission reduction effect would be a fraction of the maximum.

Second, lower natural gas prices accelerate economic activity, reduce the incentive to invest in energy-saving technologies, and lead to an aggregate expansion of the total energy system: a scale effect. All models reported greater total global primary energy consumption (6% on average) in the Abundant Gas scenario compared with the Conventional Gas scenario. All else being equal, increased energy use leads to increased CO<sub>2</sub> emissions. All models reported that the combined effect of the two forces—substitution and scale effect—does not result in a discernible reduction in emissions and, in some cases, leads to increased CO<sub>2</sub> emissions.

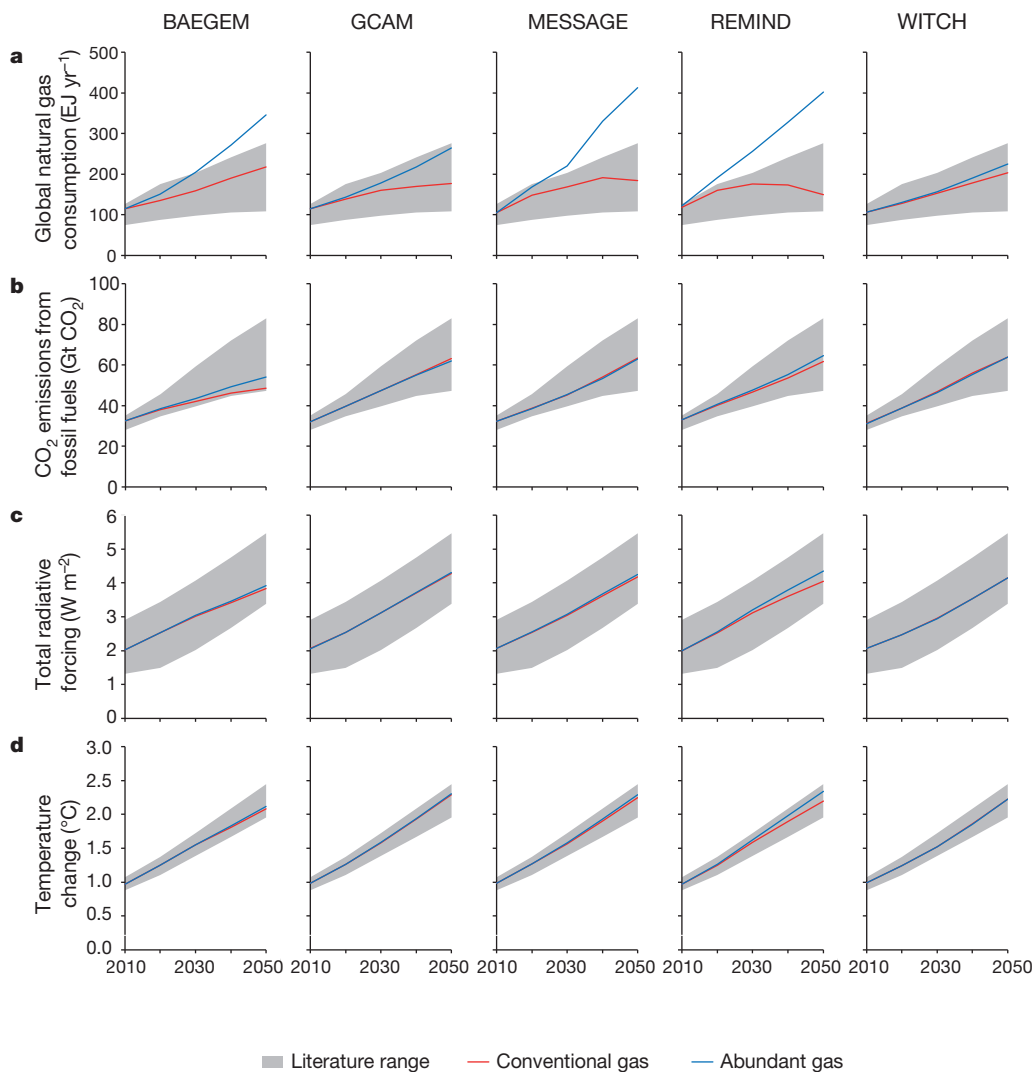
The emissions data from the models were processed through a simple climate model, MAGICC6 (Model for the Assessment of Greenhouse-Gas Induced Climate Change), to assess the combined effects of all greenhouse gases and climate forcing agents (see Methods)<sup>22</sup>. The results echoed those that were observed for CO<sub>2</sub> emissions: climate forcing and associated temperature change are not discernibly reduced under the Abundant Gas scenario (Fig. 2c, d and Fig. 3b, c). Four models that endogenously model fugitive methane emissions reported increased climate forcing with abundant gas. This is largely driven by increased forcing from fugitive methane emissions associated with increases in gas consumption. The WITCH model, with exogenously specified methane emissions, reported virtually no change in forcing (−0.3%).

Furthermore, four models reported the net change in forcing to be less than 3%. REMIND reported radiative forcing increase of 7%; more than half of that increase came from reductions in coal use and associated aerosol emissions (reduced cooling). Two other models that also simulate aerosol emissions endogenously (GCAM and MESSAGE), also reported a reduced cooling effect from aerosols, but at a smaller scale.

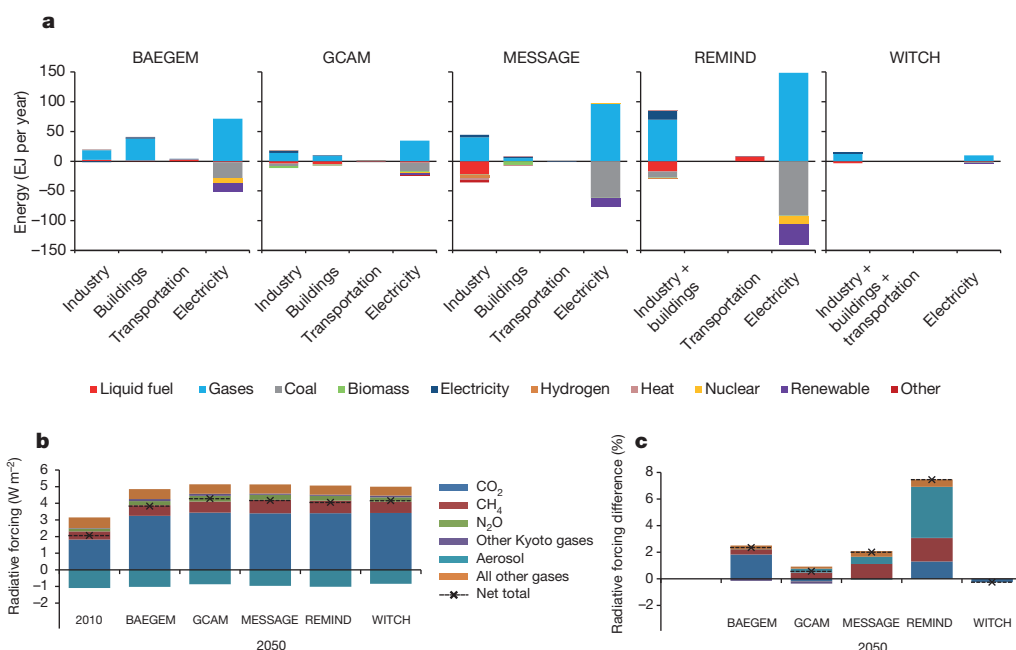
The core finding of this research is that increases in unconventional gas supply in the energy market could substantially change the global energy system over the decades ahead without producing commensurate

**Table 1 | Overview of the five modelling systems**

Model	BAEGEM	GCAM	MESSAGE	REMIND	WITCH
Full name	BAEconomics General Equilibrium Model	Global Change Assessment Model	Model for Energy Supply Systems And their General Environmental impact	Regional Model of Investments and Development	World Induced Technical Change Hybrid
Institutional steward	BAEconomics	Pacific Northwest National Laboratory (PNNL)	International Institute for Applied Systems Analysis (IIASA)	Potsdam Institute for Climate Impact Research (PIK)	Centro Euromediterraneo sui Cambiamenti Climatici (CMCC)
Location	Kingston, Australia	College Park, Maryland, USA	Laxenburg, Austria	Potsdam, Germany	Milan, Italy
Brief description	BAEGEM is a global dynamic-recursive, multi-region, multi-sector, computable general equilibrium model; it includes energy use, transformation and technology detail	GCAM is a long-term, global, dynamic-recursive, integrated assessment model of human and physical Earth systems, including 14 geopolitical and 151 land-use regions; it includes detailed technological representations for energy, land use, and the economy	MESSAGE is an integrated assessment modelling framework, combining a global (multi-region, multi-sector) systems engineering, inter-temporal optimization model, an aggregated macro-economic model, and a simple climate model	REMIND is a multi-regional, general equilibrium model of the global economy, energy, and climate systems; it includes energy supply, transformation technologies and demand details; intertemporal optimization methods solve for the equilibrium	WITCH is a multi-region, long-term, dynamic optimization, economy–energy–climate model, characterized by endogenous technological change and a game theoretic set-up with strategic interaction among regions
Climate model used for this study	MAGICC 6.0 (natively integrated with MAGICC 5.3)	MAGICC 6.0 (natively integrated with MAGICC 5.3)	MAGICC 6.0 (natively integrated with MAGICC 5.3)	MAGICC 6.0	MAGICC 6.0
Detailed description	Ref. 13	Ref. 14	Ref. 15	Ref. 16	Ref. 17



**Figure 2 | Comparison of the model results 2010–2050.** **a**, Global natural gas consumption. The relatively small difference in gas production in WITCH (11%) becomes considerably greater in the second half of the century, beyond the time scope of this study. **b**,  $\text{CO}_2$  emissions from fossil fuels. **c**, Total radiative forcing. **d**, Global mean surface temperature change (from pre-industrial average 1750–1849). The literature range is obtained from refs 29 and 30.



**Figure 3 | Global energy consumption and radiative forcing in 2050.** **a**, Differences in energy use by sector and fuel (the Abundant Gas scenario minus the Conventional Gas scenario) in 2050. One avenue for possible change in the transportation sector is through the use of gas in transportation fuel production. The MESSAGE model reports this effect at a noticeable scale (10%). **b**, Year 2010 and year 2050 composition of radiative forcing for the Conventional Gas scenario for five models. **c**, Year 2050 relative difference in radiative forcing (the Abundant Gas scenario minus the Conventional Gas scenario) for the five models. 1% difference in forcing for model average is equivalent to  $0.042 \text{ W m}^{-2}$ .



changes in emissions or climate forcing. The result stems from three effects: abundant gas substituting for all energy sources; lower energy prices increasing the scale of the energy system; and changes in non-CO<sub>2</sub> emissions. This result is potentially sensitive to a range of model assumptions.

One important assumption is that market forces are allowed to work themselves out largely unfettered. Our results would be different if policies that limit natural gas's ability to substitute for low-carbon energy were implemented on a global scale. To explore this sensitivity we recalculated the emissions assuming that abundant gas substitutes exclusively for coal. This assumption is analogous to a global clean energy standard where the capacities of carbon-free energy sources are exogenously specified. With the exception of BAEEM, the models reported CO<sub>2</sub> emission reductions between −0.1% and −6% (Extended Data Table 2). BAEEM's result of a 7% increase was driven by an overall energy expansion of 11%.

The results are also influenced by assumptions about technological change in other domains. Although the results reported here assumed changes to gas supply technology alone, oil production is experiencing similar technological advances. Extending the analysis to oil as well as gas production would not be expected to lower future CO<sub>2</sub> emissions or climate forcing because the carbon-to-energy ratio for oil is approximately 35% higher than that of natural gas.

Fugitive methane emissions associated with natural gas production, transmission, and distribution is another important factor. On the one hand, conventional estimates for natural gas methane leakage rates have been less than 2% of production<sup>23,24</sup>, and studies have shown that the leakage rate is not considerably different between conventional and unconventional sources<sup>25,26</sup>. On the other hand, other studies have reported substantially higher leakage rates<sup>7,27,28</sup>. To test the sensitivity of our results to these assumptions, we chose the highest value (7.9%; ref. 7) from a range of methane leakage rates found in the literature. We then recalculated climate forcing and found that the effect of abundant gas is to increase climate forcing by 0.2% to 12% in 2050, which is 0.5% to 5% higher than in our central scenario (Extended Data Fig. 1). In other words, the finding that abundant gas does not discernibly reduce climate forcing is consistently reported over a wide range of fugitive methane rates found in the literature. Furthermore, under high fugitive emission assumption, three models reported increased climate forcing of more than 5%.

This analysis focused solely on the potential of abundant gas to affect greenhouse gas emissions in the absence of greenhouse gas mitigation policies beyond those already in effect. The interaction between abundant natural gas and greenhouse gas mitigations policies is another issue in need of further examination<sup>9,10</sup>. Finally, we note that the global deployment of improved natural gas extraction technology carries implications not only for climate change, but also for many other important concerns including air and water quality, energy security, access to modern energy, and economic growth<sup>1,2,8</sup>.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 25 April; accepted 2 September 2014.**

**Published online 15 October 2014.**

1. Sovacool, B. K. Cornucopia or curse? Reviewing the costs and benefits of shale gas hydraulic fracturing (fracking). *Renew. Sustain. Energy Rev.* **37**, 249–264 (2014).
2. International Energy Agency. *Are We Entering a Golden Age of Gas?* (IEA Report, 2011).
3. Hultman, N., Rebois, D., Scholten, M. & Ramig, C. The greenhouse impact of unconventional gas for electricity generation. *Environ. Res. Lett.* **6**, 044008 (2011).
4. Moniz, E. J. et al. *The Future of Natural Gas: An Interdisciplinary MIT Study* (MIT, 2011); [http://mitei.mit.edu/system/files/NaturalGas\\_Report.pdf](http://mitei.mit.edu/system/files/NaturalGas_Report.pdf).
5. Brown, S., Krupnick, A. & Walls, M. *Natural Gas: a Bridge to a Low-Carbon Future* (Resource for the Future, 2009); <http://www.rff.org/RFF/Documents/RFF-IB-09-11.pdf>.
6. Levi, M. Climate consequences of natural gas as a bridge fuel. *Clim. Change* **118**, 609–623 (2013).

7. Howarth, R. W., Santoro, R. & Ingraffea, A. Methane and the greenhouse-gas footprint of natural gas from shale formations. *Clim. Change* **106**, 679–690 (2011).
8. Howarth, R. W., Ingraffea, A. & Engelder, T. Natural gas: should fracking stop? *Nature* **477**, 271–275 (2011).
9. Newell, R. G. & Raimi, D. Implications of shale gas development for climate change. *Environ. Sci. Technol.* **48**, 8360–8368 (2014).
10. Energy Modeling Forum. *Changing the game? Emissions and Market Implications of New Natural Gas Supplies* EMF Report 26 (EMF, 2013); <https://web.stanford.edu/group/emf-research/docs/emf26/Summary26.pdf>.
11. Edmonds, J. A. et al. in *Climate Change Modeling Methodology* 169–209 (Springer, 2012).
12. Rogner, H. et al. in *Global Energy Assessment* 423–512 (Cambridge Univ. Press, 2012).
13. Mi, R. & Fisher, B. S. *BAEGEM Model Documentation* (BAEconomics, 2014); <http://www.baeconomics.com.au/wp-content/uploads/2014/02/BAEGEM-documentation-21Feb14.pdf>.
14. Calvin, K. et al. *GCAM Wiki Documentation* (Pacific Northwest National Laboratory, 2011); <http://wiki.umd.edu/gcam/>.
15. Riahi, K., Grübler, A. & Nakicenovic, N. Scenarios of long-term socio-economic and environmental development under climate stabilization. *Technol. Forecast. Soc. Change* **74**, 887–935 (2007).
16. Bauer, N. et al. Global fossil energy markets and climate change mitigation—an analysis with REMIND. *Clim. Change* (in the press).
17. Bosetti, V., Carraro, C., Galeotti, M., Massetti, E. & Tavoni, M. A world induced technical change hybrid model. *Energy J.* **27**, 13–37 (2006).
18. Van Vuuren, D. P. et al. The representative concentration pathways: an overview. *Clim. Change* **109**, 5–31 (2011).
19. Intergovernmental Panel on Climate Change. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker, T. F. et al.) (Cambridge Univ. Press, 2013).
20. Bauer, N. et al. CO<sub>2</sub> emission mitigation and fossil fuel markets: dynamic and international aspects of climate policies. *Technol. Forecast. Soc. Change* (in the press).
21. Intergovernmental Panel on Climate Change *IPCC Guidelines for National Greenhouse Gas Inventories* (IPCC report, 2006); <http://www.ipcc-nggip.iges.or.jp/public/2006gl/>.
22. Meinshausen, M., Raper, S. C. B. & Wigley, T. M. L. Emulating coupled atmosphere-ocean and carbon cycle models with a simpler model, MAGICC6: Part I—model description and calibration. *Atmos. Chem. Phys.* **11**, 1417–1456 (2011).
23. Environmental Protection Agency *Inventory of Greenhouse Gas Emissions and Sinks 1990–2009* (EPA Publication 430-R-11-005, 2011); [http://www.epa.gov/climatechange/Downloads/ghgemissions/US-GHG-Inventory-2011-Complete\\_Report.pdf](http://www.epa.gov/climatechange/Downloads/ghgemissions/US-GHG-Inventory-2011-Complete_Report.pdf).
24. Venkatesh, A., Jaramillo, P., Griffin, W. M. & Matthews, H. S. Uncertainty in life cycle greenhouse gas emissions from United States natural gas end-uses and its effects on policy. *Environ. Sci. Technol.* **45**, 8182–8189 (2011).
25. Burnham, A. et al. Life-cycle greenhouse gas emissions of shale gas, natural gas, coal, and petroleum. *Environ. Sci. Technol.* **46**, 619–627 (2012).
26. Laurenzi, I. J. & Jersey, G. R. Life cycle greenhouse gas emissions and freshwater consumption of Marcellus shale gas. *Environ. Sci. Technol.* **47**, 4896–4903 (2013).
27. Miller, S. M. et al. Anthropogenic emissions of methane in the United States. *Proc. Natl Acad. Sci. USA* **110**, 20018–20022 (2013).
28. Brandt, A. et al. Methane leaks from North American natural gas systems. *Science* **343**, 733–735 (2014).
29. Kriegler, E. et al. The role of technology for achieving climate policy objectives: overview of the EMF 27 study on global technology and climate policy strategies. *Clim. Change* **123**, 353–367 (2014).
30. Calvin, K. et al. The role of Asia in mitigating climate change: results from the Asia modeling exercise. *Energy Econ.* **34**, S251–S260 (2012).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** B.F. and R.M. and their development of BAEEM were supported by BAEconomics with assistance for special applications to the resource sector from Rio Tinto (Australia). H.M., L.C., J.E. and B.P.F. were supported by the Global Technology Strategy Project. V.K., K.R. and H.R. were supported by the International Institute for Applied Systems Analysis cross-cutting project on unconventional natural gas. N.B. and J.H. were supported by funding from the German Federal Ministry of Education and Research in the project 'Economics of Climate Change'. G.M. and M.T. were supported by the Italian Ministry of Education, University and Research and the Italian Ministry of Environment, Land and Sea under the GEMINA project. We thank M. Jeong and E. Golman for research assistance. The views and opinions expressed in this paper are those of the authors alone.

**Author Contributions** H.M., J.E., L.C. and B.P.F. proposed the research design. H.R. provided the resource supply curves. H.M. and J.E. provided GCAM data and wrote the first draft of the paper. N.B. and J.H. provided REMIND data. B.F. and R.M. provided BAEEM data. V.K. and K.R. provided MESSAGE data. G.M. and M.T. provided WITCH data. All authors contributed to writing the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.M. (haewon.mcjeon@pnnl.gov).

## METHODS

**Gas supply curves.** Natural gas supply curves are harmonized across all the models. The Conventional Gas scenario represents an estimate of economically recoverable gas consistent with technology available before the shale gas revolution. This natural gas supply curve is constructed based on the synthesis of natural gas supply and geographic distribution in the GEA report<sup>12</sup>. GEA's 2010 natural gas supply curve is truncated at 11,000 EJ of cumulative global supply to represent limited conventional gas supply. No future cost reduction from technological change is assumed. This curve represents a lower bound of global natural gas supply (see green curve in Extended Data Fig. 2)

The Abundant Gas scenario represents an upper bound of global natural gas supply. This supply curve is characterized by both the global abundance of the resources and the substantially lower extraction costs. Global abundance is implemented by allowing both conventional and unconventional gas in GEA estimates<sup>12</sup> to be fully available for extraction. A cumulative supply of 39,000 EJ is assumed to be available globally.

Lower extraction cost is implemented by future technological changes reducing extraction cost over time. The extraction costs of early adopters (USA and Canada) are assumed to reduce exponentially by 1.7% per year over the course of 2011–2050. Extraction costs from all other regions are assumed to reduce exponentially by 2.0% per year over the course of 2016–2050. In all regions, the cost of extraction is reduced to half by 2050 from the GEA's 2010 estimate.

This is on the aggressive side of the cost reduction estimates found in the literature. For instance, the International Energy Agency (IEA)'s 'Golden Age of Gas' scenario<sup>2</sup> shows a 23% reduction by 2035; Newell and Raimi<sup>9</sup> assumes a 45% reduction by 2040; the EMF 26 model comparison exercise's 'High Shale' scenario<sup>10</sup> assumes a 21% reduction by 2035; and finally, the original GEA scenarios<sup>12</sup> assumed a 33% reduction by 2050. Although direct comparison is difficult owing to each scenario's differences in design, measurement, and time frame, our aggressive assumption is intended to represent a lower bound for future gas production costs.

The cost reduction assumption is also more aggressive than that of relatively mature low-carbon energy sources such as nuclear power plants, but not necessarily more aggressive compared to that of immature technologies such as solar photovoltaics (see Extended Data Table 1 for cost reduction comparison with other energy sources).

When the abundance in quantity and reduction in production cost are combined, the Abundant Gas supply curve allows 31,000 EJ of cumulative natural gas production at \$3 per GJ or less by 2050. Extended Data Fig. 2 shows the reduction in cost from 2010 (red curve) to 2050 (blue curve). These production costs are different from the actual prices in the market place. The costs do not include taxes or royalties, nor do they include external environmental or social costs associated with gas production<sup>12</sup>.

GCAM and MESSAGE models further tested sensitivity to the magnitude of production cost reductions. GCAM is a model relatively less sensitive to Abundant Gas supply, while MESSAGE is a model relatively more sensitive to it. The two models projected a total of five natural gas supply scenarios: (1) the Conventional Gas scenario; (2) the Abundant Gas scenario (at the standard 50% cost reduction); (3) the Abundant Gas scenario (with the high cost reduction of 75%); (4) the Abundant Gas scenario (with the low cost reduction of 25%); (5) the Abundant Gas scenario (with the zero cost reduction, abundant in quantity only).

The results are shown in Extended Data Fig. 3. Collectively, these scenarios cover a wide range of cost reduction found in the literature. Our core finding from the main analysis is found to be consistent with results from this sensitivity. In all cases considered, we found that more abundant natural gas could substantially change the global energy system over the decades ahead without producing commensurate changes in emissions or climate forcing.

GCAM reported +13% to +82% additional natural gas consumption in 2050, while the change in CO<sub>2</sub> emissions is found to be in the −0.9% to −2.0% range and the change in radiative forcing is found to be in the +0.3% to +1.1% range. MESSAGE reported +56% to +170% additional natural gas consumption in 2050, while its change in CO<sub>2</sub> emissions is found to be in the −1.0% to +0.6% range and its change in radiative forcing is found to be in the +0.7% to +3.4% range.

Just as in the main analysis, the models did not agree on the direction of the impact on CO<sub>2</sub> emissions. GCAM consistently reported lower CO<sub>2</sub> emissions with respect to lower cost assumptions. MESSAGE reported that the CO<sub>2</sub> emissions increase at the high end of the cost reduction range. However, these changes are very small, with a magnitude less than 2% of the total emissions. Once we consider the combined effect of all greenhouse gases, the two models consistently agree on the direction of the change: the lower the natural gas production cost, the higher the total radiative forcing and associated temperature change. Our main finding that increased use of abundant gas does not produce a discernible reduction effect on climate forcing is found to be consistent across the range of cost reduction sensitivities.

**Main analysis.** The main analysis presented in the paper may be sensitive to a range of model assumptions. We reported two of the core sensitivities in the main text. Here we describe detailed methodologies for the main analysis and the sensitivity analyses.

The main analysis follows a standard method for IAM study on baseline scenarios. The five models are explicitly designed to project the future emissions trajectory under various assumptions about the energy system and the economy. Representing the energy and economic system in an abstract structure, IAMs provide a simulation method of conducting an analysis when a 'controlled experiment' in the strict sense is not possible. Similar to a controlled experiment, our numerical experiment keeps all other parameters constant and varies only the natural gas supply curve. We then simulate the effect of market forces on the energy system evolution through 2050. From the two simulations that differ only in terms of natural gas supply, we report the differences in the output variables, such as energy system composition, emissions, and climate forcing. Such differences in output variables are directly attributable to the differences in input variables.

To closely replicate the human system dynamics, each model calibrates its parameters to the observed data in the historical years. The data used for calibration is reported in the model descriptions. Calibrated parameters include technological parameters such as energy production efficiency and emissions intensity, economic parameters such as price elasticities and income elasticities, and non-market parameters such as regional preferences for specific fuel type or preferences for a specific mode of travel. Projecting into the future, some parameters are assumed to improve over time (for example, energy production efficiency) and others are assumed to be constant (for example, social preferences for a specific mode of travel are assumed to be constant).

As this is a baseline scenarios study, we did not assume any explicit climate change mitigation policies. This study addresses the following question: if there are no new policies to mitigate climate change, does increased use of abundant natural gas reduce greenhouse gas emissions? However, although no economy-wide climate change mitigation policy has been currently implemented, the currently existing policies that have been implemented in the past would affect the parameters calibrated to historical observations. For instance, the Corporate Average Fuel Economy (CAFE) standard in the USA has been enforced since 1978 to increase the fuel economy of cars and decrease fuel consumption. Although this was not explicitly intended as a climate change mitigation policy, it has had the side-effect of reduced emissions per distance travelled. This side-effect would be implied in the calibration process and propagate forward into the future. As a result, the future projection of emissions would be lower than they would otherwise have been without the CAFE standard in effect.

Similarly, any energy policies that were enforced before the calibration periods would affect the calibration parameters. These include Renewable Fuel Standard policies that mandate the ratio of biofuels in gasoline, building energy standards that mandate the minimum efficiency levels of building shells, and renewable or fossil fuel energy subsidies. These policies affect the implied preference for certain energy sources or efficient equipment. However, we do not include the policies that are proposed, but not currently in effect. For studies that do include proposed policies in their scenarios, see refs 31 and 32. Next, we describe one sensitivity analysis that explicitly represents a future energy policy that is currently not included in the calibration.

**Abundant gas exclusively substitutes coal scenario.** In our main analysis models allow natural gas to substitute not only for coal, but also for a range of energy sources including solar, wind, nuclear, and bioenergy. These substitutions are driven by the economic competitiveness of each fuel type. However, it is also possible to imagine a policy architecture in which a normative policy protects low-carbon energy sources, thus effectively forcing additional natural gas to exclusively substitute coal. Under such restrictions, we expect overall CO<sub>2</sub> emissions to decrease. To estimate the magnitude of the sensitivity to the substitution restriction, we assume a future where low-carbon energy sources are protected by a globally enforced Clean Energy Standard. In this scenario, natural gas is assumed to exclusively substitute for coal.

First, we assume that the low-carbon energy quantity is fixed at the same level as the Conventional Gas scenario. Then, we calculate the quantity difference in low-carbon energy between the Conventional Gas and Abundant Gas scenarios; this is the amount of low-carbon energy that would be protected under the policy. To keep the scale of energy system unchanged, we assume the same amount of coal is instead substituted by additional gas. The total amount of gas consumption remains unchanged. We then apply the emissions factors from Extended Data Table 3 and recalculate the additional emission reduction.

With the exception of BAEGEM, all models show that the 'coal substitution only' assumption results in emission reductions in 2050, ranging from −0.1% (WITCH) to −5.9% (MESSAGE). See Extended Data Table 2 for the range of values. Compared to the emissions changes in the main analysis, in which the majority of the

models showed positive emissions increase, the 'coal substitution only' scenario shows that under certain policy conditions, abundant gas could help reduce CO<sub>2</sub> emissions.

In the case of BAEDEM, the lower gas prices accelerate economic activity, such that the overall energy system is 11% larger in the Abundant Gas scenario. The 'coal substitution only' assumption does reduce the average emission intensity of the energy system, but the energy system expansion effect still dominates, such that the total emission is still larger than in the Conventional Gas scenario.

**High fugitive methane emissions scenario.** The fugitive methane emission rate is subject to large uncertainty. The rates used in the five models all fall within the range 0.3–0.6 kg of CH<sub>4</sub> per GJ (Extended Data Table 3). These values are similar to the values reported in conventional literature<sup>23,24,33</sup>. However, some recent literature suggests that the fugitive methane rate may be substantially higher by up to a factor of four<sup>7,27,28</sup>. To test our results' sensitivity to high fugitive methane rates, we select the upper bound of fugitive methane estimates (7.9%) found in the literature<sup>7</sup> and re-estimate the climate forcing.

We start from the original emission trajectories from each model. Then, while keeping all else equal, we recalculate the methane emission trajectory by applying the high fugitive methane emission rate to the natural gas use. These modified emission trajectories are then reprocessed through the common climate model MAGICC6. With the high fugitive methane assumptions, the abundant gas increases the total anthropogenic radiative forcing by 0.2% to 12% in 2050, which is 0.5% to 5% points higher than under the standard assumptions. A full comparison is shown in Extended Data Table 4.

**IAMs of energy–economy–climate systems.** The five models that are used in this study are members of a class of models referred to IAMs. IAMs in general encompass the broad suite of human and natural Earth systems including the economy, energy, agriculture, land-use, land cover, and biogeophysical processes from carbon and hydrologic cycles, the atmosphere, oceans and climate<sup>11</sup>. The five models employed in this study are well equipped to assess the impact of abundant natural gas on climate forcing. Each contains a state-of-the-art energy–economy systems model coupled to a simple climate model.

Each of the five models represents energy and economic systems differently. Below, we provide a general description of the strengths and limitations the five models bring to the issue of assessing the impact of abundant natural gas on climate forcing. This is followed by more detailed descriptions of the five models. The IAMs of energy–economy–climate systems employed in this study bring a number of strengths to the issue of the global long-term climate forcing implications of abundant natural gas. In general, the models were designed to address precisely the kind of problem we explore in this paper. They have the appropriate geographic, temporal, and sectoral coverage.

All of the models explicitly represent processes that start from the extraction of primary energy (exhaustible fossil fuels and renewable energy) to energy transformation (for example, liquid fuel refineries and power generation) to end-use services (buildings, industry, and transport). The models feature explicit representation of energy markets with price-responsive demand and supply for coal, oil, and gas as well as low-carbon energy sources<sup>20,34</sup>. The flexibility and interdependence of energy markets are crucial for the present study because these features determine the degree to which additional natural gas is consumed and by how much this reduces the demand for other fuels. All five models employ a standard economic paradigm in their representation of energy markets. Price is the principal force determining and equilibrating the supply of and demand for different fuels.

IAMs vary in a number of important ways. While all of the models in this study have explicit representations of both the economic system and the energy system, they vary in terms of their relative emphasis on representing the details of the two interlinked systems. Model structures that emphasize economic interactions across all sectors of the industry are particularly strong for examining how changes in one industry propagate through the whole economy. These models are also strong in examining changes in international trade patterns due to region-specific changes in industrial structure. BAEDEM<sup>13</sup> is one example of this type of model with 25 explicit sectors of the economy each consuming a bundle of energy sources, where the share of the bundle is determined by the relative prices.

In contrast, GCAM<sup>14</sup> and MESSAGE<sup>35</sup> place greater emphasis on representing the details of the physical energy system. They contain detailed representations of key energy systems and technology options for producing, transforming and using energy, while adopting more aggregate representations of the broader economy. GCAM and MESSAGE have more than 100 different energy supply and conversion technology representations. This approach is more rigid in the ability to substitute between the factors of production, namely capital/labour inputs and energy inputs, compared to the approach used by models such as BAEDEM. However, it can better capture the physical details of individual services provided in the end-use sectors, such as ton-kilometres of freight or GJ of residential heating. This modelling approach is particularly strong for in-depth analysis of a specific energy technology and tracking the physical flows of energy goods and services.

REMIND<sup>36</sup> and WITCH<sup>17</sup> lie between BAEDEM on the one end and GCAM and MESSAGE on the other. These models have more detail in their economic structure and less energy system detail than the latter, but more energy system detail and less economic system detail than the former.

Another domain in which the models vary is their assumption about the knowledge and behaviour of their economic agents. Intertemporal optimization models represent economic agents that maximize their economic utility over the model time horizon. The agents are assumed to know future price changes with certainty and hence the resulting model solution is economically optimal for each agent. MESSAGE, REMIND, and WITCH employ this method.

On the other hand, GCAM and BAEDEM takes a more descriptive approach where the economic agents are assumed to not know the future price changes, but rather make production and consumption decisions based only on the information available to them at any given time. This approach provides insights about how the system might be expected to evolve under imperfect information about the future. However, the resulting equilibrium trajectory may not be economically optimal.

**Modelling paradigm for IAMs.** The variety in modelling approaches used by the five models in this exercise strengthens this analysis. This diversity guards against a result that is the product of an individual model's idiosyncratic behaviour. To the extent that models employ different methodologies and get qualitatively similar results, we have greater confidence in the result. That said, all models subscribe to the standard economics paradigm. Other modelling paradigms exist, such as adaptive agent models, system dynamics models, or infrastructure models. We have not tested these modelling paradigms in this paper. As such, the models do not span the full range of all possible methods that could potentially be employed to assess the impact of more abundant natural gas for climate forcing.

One limitation of the models employed in this study is that they do not model explicit locations of physical infrastructure. The geographic locations of present and future natural gas pipelines and liquefied natural gas terminals are modelled at coarse international resolution, and do not take into account detailed local information that shapes decisions about which facilities are deployed, and where and how they are connected to the broader system. Geographically resolved infrastructure models can potentially include this level of detail. However, infrastructure models are relatively static in nature and are therefore generally not employed to model the global energy system's evolution over multiple decades into the future.

Also, the five models are built on the foundation of the standard economic paradigm, and they do not, for instance, employ an adaptive agent modelling approach or systems dynamics approach. Future research could employ a broader suite of modelling methods to shed further light on the implications of abundant natural gas for climate forcing, and examine whether other modelling approaches would yield a qualitatively different result.

**Representing the policy environment in IAMs.** The energy policy environment exerts a strong influence on energy production and use and thereby on climate forcing. Our default assumption is that no new policies and measures are introduced after the calibration period. Alternative assumptions can produce different results for energy, for the economy and for climate forcing. We tested one energy policy that can potentially change the results. We found that exogenously specifying the quantity of low-carbon energy sources and forcing natural gas to substitute exclusively for coal results in emissions being reduced in the models.

Other policies, such as carbon tax, cap-and-trade, or natural-gas vehicle subsidies, could alter our results. The climate implications of abundant gas under climate policies are of great interest. However, the issue is sufficient in scope and depth to require its own future research, and hence is not addressed here.

**Baseline assumptions in IAMs.** Finally, we point out that the numerical simulations of the effects of increased natural gas availability that we have performed for this paper are all based on each model's native reference scenario assumptions. Those scenarios are each developed by the modelling teams themselves and no attempt was made to harmonize assumptions other than natural gas supply curves. This was intentionally done to increase the variety of conditions against which the implications of abundant gas would be assessed.

Specifying different exogenous assumptions would produce different results. Some perturbations have well established consequences common to all of the models. For example, higher population growth or higher rates of economic productivity growth increases the scale of the energy system overall. Perturbations in the assumed rate at which technological change that occurs in low-carbon technologies would change the future emission intensity of the energy system.

While we have not attempted to explore the sensitivity of our results to variation in those assumptions, the five models' native assumptions cover a wide uncertainty range consistent with the large majority of the literature<sup>37</sup>, but they do not cover the extremes found in the literature (see Extended Data Fig. 4). Examining the four principal components of model projections (population, gross domestic product (GDP), energy consumption, and CO<sub>2</sub> emissions), we observe that the five models



cover most of the 10th to 90th percentile range for population and GDP in 2050. The projected ranges for energy consumptions and CO<sub>2</sub> emissions are narrower, but cover most of the 25th to 75th percentile range. To the extent that the five models report qualitatively similar results despite the large variations in baseline assumptions, we have greater confidence in the results.

**Uncertainties in IAMs.** Uncertainty attends every element of the modelling process. There are two major elements of uncertainty in the modelling system: model structure and model input assumptions (future population, economic activity, technology and policy). Methods have been developed to address each of these uncertain elements. Uncertainty in model input assumptions are addressed in a variety of ways ranging from simple sensitivity analysis to formal uncertainty quantification analysis. Uncertainty in model structure is more difficult and model intercomparison is an important tool for exploring this source of uncertainty.

The formal Monte Carlo analysis<sup>38,39</sup> employs a single model to process numerous samples of input variables, such as fossil fuel supply, economic growth, technological learning, and so on. The distribution of the output shows the range of uncertainty associated with the model results. For simple models, uncertainties associated with all model inputs can be examined. However, with the growing complexity of IAMs, this method is becoming increasingly difficult to implement.

Sensitivity analyses identify key variables of interest and examine each model's response to variation in those input values. We have identified three such variables and performed sensitivity analyses. We have also examined sensitivity to energy policy, specifically sensitivity to low-carbon energy policy. In this latter case we found that under a global low-carbon energy protection policy, the availability of more abundant gas can reduce climate forcing. We also explored the sensitivity of our results to the rates of fugitive methane emissions and found that with high fugitive methane emission rates, more abundant gas can discernibly increase climate forcing. Finally, we explored the sensitivity of model outputs to natural gas supply characterizations, specifically production cost, and found that our results were consistent across a wide range of natural gas supply assumptions.

Model intercomparison projects (MIPs), where a number of models simulate a common set of scenarios, are the primary method employed to explore the implications of variation in model structure, although they are also used to structure sensitivity and scenario analysis. The Energy Modelling Forum (EMF)<sup>10</sup> has been conducting MIPs of energy–economy models since 1977. The MIPs conducted by EMF are analogous to the Coupled Model Intercomparison Project (CMIP)<sup>40</sup> of the climate modelling community, where a larger number of climate models project the future climate and assess the distribution of the projection. MIPs can be thought of as the modelling equivalent of scientific hypothesis testing using different methods. If a number of models with heterogeneous architecture reach a common conclusion, we can have greater confidence in that conclusion.

This study is an example of a MIP with a small number of models. The comparison of the results across the models shows large uncertainty. The uncertainty is especially large in the future level of natural gas consumption, and consequently the use of competing energies, such as solar photovoltaics and nuclear power plants. The uncertainty is also present in the size of the impact of abundant gas on the emissions. These results are highly dependent on model architecture and the implied flexibility of fuel-switching. However, the models all agree on the most potent conclusion: increased supply of abundant gas does not discernibly reduce either CO<sub>2</sub> emissions or climate forcing. Some models report a discernible increase in emissions or climate forcing, and others report negligible change. But none of the models report more than a 2% reduction in emissions or climate forcing. This qualitative agreement across five heterogeneous models in this exercise gives greater confidence in the conclusion.

**Overview of the BAEDEM model.** BAEDEM<sup>13</sup> is a recursively dynamic computable general equilibrium model of the world economy. For each one-year time step, BAEDEM simulates the interrelationships between economic growth, flows of international trade and investments, constraints on natural resources and production factors, greenhouse gas emissions and climate change policies.

The central core of BAEDEM is built on the familiar approaches of the GTAP model<sup>41</sup>, with the household consumption behaviour and the producer behaviour represented separately by a constant difference of elasticities function and a nesting of Leontief, constant elasticity of substitution (CES) and constant ratios of elasticity of substitution, homothetic (CRESH) functions<sup>42</sup>.

BAEDEM is written in GEMPACK<sup>43</sup>. The full model code is complemented by four interlinked modules: (1) the government module; (2) the technology mix module; (3) the energy module; and (4) the greenhouse gas emissions module. The model is ideally suited to analysing domestic and international energy-related policies, and the impacts of economic shocks.

The BAEDEM database is derived from a number of sources. The global social accounting matrix is derived from the GTAP version 8 database<sup>44</sup> with a base year of 2007. To enhance the capability of modelling individual commodities, the number of commodity groups in BAEDEM has been expanded from 57 in the GTAP version

8 database to 72. The disaggregated commodities include black thermal coal, brown coal, coking coal, iron ore, bauxite, copper ore, gold, uranium, titanium, zirconium, coke, nuclear fuel, alumina, copper, aluminium and liquefied natural gas.

The emissions database covers all Kyoto gases and is sourced from the International Energy Agency<sup>45,46</sup>, the United Nations Framework Convention on Climate Change<sup>47</sup> and the US Environmental Protection Agency<sup>48</sup>. The data in the government module are sourced from Global Insight while the data in the technology mix and energy modules are sourced from the IEA.

The global temperature rise, total radiative forcing and the atmospheric concentration of carbon dioxide can be calculated from the BAEDEM results by linking to MAGICC<sup>49</sup>, with climate sensitivity set to three degrees Celsius. BAEDEM natively links to MAGICC 5.3, but for this study we have used MAGICC 6.0 for latest scientific knowledge and consistency across models.

Supplementary Fig. 1 and Supplementary Table 1 provide an overview of the other key features of BAEDEM.

**Modelling energy commodities in BAEDEM.** The energy module tracks the production of primary and secondary energy, and the consumption of final energy by government, households and firms. Changes in production volume over the projection period are driven by global demand growth, which in turn is determined by real GDP growth, and changes in prices, consumption preference, market structure, sector productivity and market structure.

The government demand for each commodity is derived from a Cobb–Douglas function nested with Armington composites of commodities supplied by domestic and foreign sources. The household demand for each commodity is determined by the demand of a representative household and the growth in population. At the first level, the representative household chooses quantities of non-energy commodities and an energy composite (that is, coal, gas, refined petroleum product, electricity and heat) to maximize a utility function, given a budget constraint. At the next level, the representative household chooses quantities of energy commodities to minimize the cost of consuming the energy composite in the previous level. The purpose of this two-level demand system is to reflect better the substitutability between energy commodities with a more flexible substitution system.

Demands for energy commodities in each production sector are derived from a nesting of Leontief, CES and CRESH functions. At the first level, a Leontief technology links the input of factor–energy composite to the industry output level. At the second level, it is a CES cost minimization problem searching for an optimal combination of energy and factor composites where energy commodities and primary factors (that is, capital, labour, land and natural resources) are substitutable, but not perfectly so. For land and natural resource-intensive industries (that is, crops, livestock, coal, oil, and gas), a CES structure with imperfect substitutability ensures that constraints on land and natural resource or more intensive use of capital and labour under finite natural resources can be modelled properly in BAEDEM. At the third level, another cost minimization problem is specified, but here it searches for an optimal combination of energy commodities under a CRESH production function.

Electricity supply from various technologies is modelled inside the technology mix module. The ‘technology bundle’ approach ensures that electricity output can be produced from a bundle of individually identified generation technologies and that each technology uses a different mix of inputs. The purpose of integrating a bottom-up modelling approach for the electricity sector into BAEDEM is to represent better the technology-specific detail of the sector while retaining the benefits of the top-down interactions modelled in BAEDEM. In this application, the electricity output is the sum of nine technologies: coal; oil; gas; nuclear; hydro; wind; solar; biomass; and others.

The substitution possibilities between electricity technologies in BAEDEM are governed by a CRESH aggregation function. CRESH is a generalization of CES and allows elasticity of substitutions to vary between its elements. In other words, certain technologies identified in the framework can be assumed to be more substitutable than others. The use of the family of CRESH aggregation functions allows for the fact that electricity, which is a homogenous output, can be generated in an economy simultaneously from different technologies with different production costs.

**Modelling greenhouse gas emissions in BAEDEM.** The greenhouse gas module tracks Kyoto gas emissions (for CO<sub>2</sub>, CH<sub>4</sub>, N<sub>2</sub>O, HFCs, PFCs and SF<sub>6</sub>) over the course of production, transformation, consumption, and combustion. For each time step, emissions pathways of Kyoto gases are derived from the quantities of these economic activities and changes in emission factors. The projections of radiative forcing agents other than Kyoto gases are selected from emission scenarios in MAGICC according to modelling criteria, assumptions and applications. Supplementary Table 2 provides the list of the gases and the data sources for assigning emissions coefficients for each sector in BAEDEM.

BAEDEM assumes the constant proportionality of emissions with respect to the quantity of fossil fuel combusted over time. The disaggregated CO<sub>2</sub> emissions for the base year is derived from the GTAP 8.0 database with adjustments to ensure

that aggregate combustion emissions at country level are consistent with the IEA combustion emission database<sup>50</sup>.

Non-combustion emissions, such as fugitive emissions from fossil fuel mining, enteric fermentation in livestock production and chemical transformation in manufacturing processes, are assumed to change in proportion to their production levels adjusted by EMF21 marginal abatement curves<sup>51</sup>. The use of marginal abatement curves in the module allows a gradual reduction of non-combustion emissions per unit of output with additional reduction opportunities when carbon price increases. The disaggregated non-CO<sub>2</sub> emissions for the base year is derived from the US Environmental Protection Agency database<sup>48</sup> and the GTAP 7.0 database with adjustments to ensure that aggregated non-CO<sub>2</sub> emissions are consistent with the IEA non-CO<sub>2</sub> emissions database<sup>46</sup>.

**Overview of the GCAM model.** GCAM is a global integrated assessment model of energy, economy, land-use, and climate. GCAM originates from the Edmonds and Reilly model<sup>52–56</sup>. In this paper, we use the standard release of GCAM 3.1 with the natural gas system specifically modified to reflect the common assumptions on natural gas. GCAM is an open-source model<sup>14</sup> primarily developed and maintained at the Joint Global Change Research Institute. The full documentation of the model is available at the GCAM wiki<sup>14</sup>, and the following description is a summary of the wiki documentation.

GCAM is a long-term global model with particular emphasis on the representation of human dimensions of the Earth system. GCAM integrates representations of the global economy, energy systems, agriculture and land use, with representation of terrestrial and ocean carbon cycles, and a suite of coupled gas-cycle and climate models (Supplementary Fig. 2).

The climate and physical atmosphere in GCAM is represented by MAGICC<sup>22</sup>. The emission trajectories of greenhouse gases are modelled in GCAM's energy and land-use components. GCAM is natively integrated with MAGICC 5.3, but for this study we have used MAGICC 6.0 for the latest scientific knowledge and consistency across models.

The global economy of GCAM is represented in 14 geopolitical regions, explicitly linked through international trade in energy commodities, agricultural and forest products, and other goods such as emissions permits. The scale of economic activity is driven by population size, age and gender, and labour productivity, which determine economic output in each region. The energy and land-use market equilibrium is established in each period by solving for a set of market-clearing prices for all energy and agricultural good markets. This equilibrium is dynamic-recursively solved for every five years in the period 2005–2100. Supplementary Table 3 provides an overview of the other key features of GCAM.

**Modelling energy system and natural gas in GCAM.** In GCAM, the energy system represents processes of energy resource extraction, transformation, and delivery, ultimately producing services demanded by end users. Resources are classified as either depletable or renewable; in either case, the extraction costs of a given resource are assumed to increase as economically attractive resources are employed, but are also subject to technological progress which can lower extraction costs for a given resource grade. In each time period, the market prices of energy goods and services, including fossil fuel resources, are determined within the market equilibrium.

Fossil fuel energy is produced from a graded, regionally disaggregated depletable resource base. Renewable energy forms are also disaggregated by region and resource grade; however, by their nature, the resource is not consumed by use. Primary energy forms can be transformed into final energy products, including electricity, processed gas products, refined liquids, and so on.

Energy transformation sectors convert resources initially into fuels consumed by other energy transformation sectors, and ultimately into goods and services consumed by end users. Multiple technologies compete for market share; shares are allocated among competing technologies using a logit choice formulation<sup>57</sup>. The cost of a technology in any period depends on two key exogenous input parameters—the non-energy cost and the efficiency of energy transformation—as well as the prices of the fuels it consumes. The non-energy cost represents all fixed and variable costs incurred over the lifetime of the equipment (except for fuel costs), expressed per unit of output. For example, a gas-fired electricity plant incurs a range of costs associated with construction (a capital cost) and annual operations and maintenance. The efficiency of a technology determines the amount of fuel required to produce each unit of output. The prices of fuels are calculated endogenously in each time period based on supplies, demands, and resource depletion. The depletion of economically available energy resources is explicitly tracked throughout the modelling period.

The natural gas resource supply curves for the two scenarios are based on synthesis by the GEA<sup>12</sup>, as described above. In GCAM, natural gas can be used for direct combustion in the end-use sectors or converted into other energy forms, such as electricity hydrogen or refined liquids, before being consumed in the end-use sectors. Direct combustion and conversion to other forms both result in CO<sub>2</sub> emissions. The physical quantity of carbon is preserved throughout the energy system process. Once natural gas is extracted, the carbon in the fuel is either emitted or sequestered. Non-CO<sub>2</sub>

emissions are tracked separately. The next subsection describes the treatment of non-CO<sub>2</sub> emissions in detail.

**Modelling greenhouse gas emissions in GCAM.** GCAM tracks 16 different greenhouse gases, aerosols and short-lived species. Supplementary Table 4 provides the list of the gases and the data sources for assigning emission coefficients for each sector in GCAM.

Fossil fuel CO<sub>2</sub> emissions are modelled according to the following method: (1) The total emission in the base year is calibrated to the Carbon Dioxide Information Analysis Center database<sup>58</sup>. (2) The fossil fuel consumption in the base year is calibrated to the IEA's Energy Balances Database<sup>59,60</sup>. (3) The average emission coefficients are derived from the ratio of the total emission and the total fuel consumption for each fuel (coal, oil, and gas). (4) These emission coefficients are applied to each sector in the base year. (5) For future periods, GCAM solves for market shares of each fuel in each sector, and the emissions are calculated to be the product of emission coefficients and the fuel consumption in each sector.

Non-CO<sub>2</sub> gases in the energy system are calculated according to the following method: (1) The total emission for each gas in the base year is calibrated to the RCP data<sup>18,61</sup>. (2) Emissions by each sector in GCAM are compiled from the databases listed in Supplementary Table 4. (3) The individual emission coefficients for the base year are calculated by scaling the individual sector emissions to match their sum to the total emissions. (4) For future periods, GCAM solves for market shares of each technology in each sector, and the emissions are calculated by the product of emission coefficients and the technology usage level in each sector. (5) Future emission coefficients are assumed to improve over time with economic growth based on Energy Modelling Forum Study 21<sup>48,51</sup>.

Extended Data Table 3 shows the calculated emissions coefficients of CO<sub>2</sub> and CH<sub>4</sub> emissions for each fossil fuel. Fugitive CH<sub>4</sub> emission for natural gas is modified in the sensitivity analysis to reflect the wide range of estimates in the literature.

**Overview of the MESSAGE model.** MESSAGE<sup>35,62,63</sup> is a linear-programming systems engineering optimization model used for medium- to long-term energy system planning and policy analysis. The model minimizes total discounted energy system costs, and provides information on the utilization of domestic resources, energy imports and exports and trade-related monetary flows, investment requirements, the types of production or conversion technologies selected (technology substitution), pollutant and greenhouse gas emissions, and inter-fuel substitution processes, as well as temporal trajectories for primary, secondary, final, and useful energy.

MESSAGE stands at the core of the IIASA integrated assessment framework<sup>15</sup>, which combines a blend of different models to represent the global economy and the interactions between energy, agriculture, and forest sectors and their implications for greenhouse gas emissions and associated climate responses.

MESSAGE is linked to the macro-economic model MACRO for assessing economic feedbacks and price-induced changes of energy demand<sup>62</sup>. In the form used here, MACRO has its roots in a long series of models by Manne and Richels, the latest of which is MERGE 5.1<sup>64</sup>. MACRO's objective function is the total discounted utility of a single representative producer–consumer (for each of its 11 world regions). The maximization of this utility function determines a sequence of optimal savings, investment, and consumption decisions. In turn, savings and investment determine the capital stock. The capital stock, available labour, and energy inputs determine the total output of an economy according to a nested CES production function. Energy demand in six categories (industry electric and thermal, residential electric and thermal, transport and non-energy use) is determined within the model, consistent with the development of energy prices and the energy intensity of GDP. When MACRO is linked to MESSAGE, internally consistent projections of GDP and energy demand are calculated in an iterative fashion that takes price-induced changes of demand and GDP into account. This is achieved through iterations between the two models, in which demand, energy system costs and energy prices are exchanged until the solution of both models converge. For details of the iterative model linkage, see ref. 62.

In addition to the energy sector, MESSAGE represents the greenhouse gas emissions from land-use changes in the agricultural and forest sector. For the calculation of physical climate responses, MESSAGE is coupled with MAGICC<sup>55</sup>. MESSAGE is natively integrated with MAGICC 5.3, but for this study we have used MAGICC 6.0 for the latest scientific knowledge and consistency across models. Supplementary Table 5 provides an overview of the other key features of MESSAGE.

**Modelling energy system and natural gas in MESSAGE.** A typical model application is constructed by specifying performance characteristics of a set of technologies and defining a Reference Energy System that includes energy technologies and flows along the entire energy chain. In the course of a model run MESSAGE determines how much of the available technologies and resources is actually used to satisfy a particular end-use demand, subject to various constraints, while minimizing total discounted energy system costs. A simplified illustration of the MESSAGE Reference Energy System is shown in Supplementary Fig. 3.

The representation of the energy system includes explicit tracking of the long-lived energy infrastructure by vintage, which allows for consideration of the timing of technology diffusion and substitution, the inertia of the system for replacing existing facilities with new generation systems, clustering effects (technological interdependence) and possible phenomena of increasing returns (that is, the more a technology is applied the more it improves and widens its market potentials). Combined, these factors can lead to 'lock-in' effects<sup>66,67</sup> and path dependency (change occurs in a persistent direction based on an accumulation of past decisions). As a result, technological change can go in multiple directions, but once change is initiated in a particular direction, it becomes increasingly difficult to change its course.

Important inputs for MESSAGE are technology costs and technology performance parameters. For the scenarios included in this paper, technical, economic and environmental parameters for over 100 energy technologies are specified explicitly in the model. Costs of technologies are assumed to decrease over time as experience (measured as a function of cumulative output) is gained. Assumptions for the main energy conversion technologies are summarized in ref. 68. The regional energy costs are based on IEA<sup>69</sup>. For carbon capture and storage technologies, the power sector applications are based on ref. 70 and the liquid conversion processes are based on refs 71–73. Biomass technology costs are based on ref. 71. For the evolution of technology costs over time we adopt the assumptions of the GEA-Mix scenario of the Global Energy Assessment<sup>74</sup>.

Fossil fuel resource estimates and potentials for renewable energy are another important set of input parameters. For fossil fuel availability the model distinguishes between conventional and unconventional resources for different categories of oil, gas, or coal occurrences<sup>75</sup>. With regard to volumes for coal and oil we mainly follow the quantitative assumptions adopted by the GEA<sup>76</sup>. Resource assumptions for natural gas are different across the scenarios in this paper and were specifically updated from ref. 12 to represent 14 different occurrences of natural gas for each of the 11 MESSAGE regions. Energy losses (own use) of natural gas extraction are modelled explicitly and range from close to zero up to 25% of the extracted gas, depending on the type of natural gas occurrence. Fugitive CH<sub>4</sub> emissions from natural gas extraction are assumed to be between close to zero to 5% of the extracted natural gas and increase for unconventional gas resources. Assumptions about energy losses and fugitive emissions are based on ref. 77. For renewable energy resource potentials we rely on spatially explicit analysis of resource availability and adopt the assumptions discussed in ref. 68.

Representation of natural gas infrastructure in the MESSAGE model comprises explicit technologies for extraction, transmission and distribution, trade, conversion and use of natural gas in appliances of various service sectors. Main energy conversion technologies include various types of power generation technologies, heat generation (including combined heat and power facilities), hydrogen generation, and gas-to-liquid supply chains. Intra-regional trade options include piped gas as well as liquefied petroleum gas. Natural gas consumption of end-use appliances are modelled at the level of three main energy end-use sectors, including residential and commercial, industry and the transport sector. CO<sub>2</sub> emissions are modelled along the conversion chain and are either vented to the atmosphere or sequestered underground in the case of carbon capture and storage.

**Modelling greenhouse gas emissions and CH<sub>4</sub> in MESSAGE.** In addition to CO<sub>2</sub> emissions, the MESSAGE model considers the full basket of non-CO<sub>2</sub> greenhouse gases (CH<sub>4</sub>, N<sub>2</sub>O, and F-gases) as well as emissions from other radiatively active substances from the energy, industrial and non-energy sectors of the economy (disaggregated at each of the model's eleven regions). These include particulate matter (PM<sub>2.5</sub>), sulphur dioxide (SO<sub>2</sub>), nitrogen oxides (NO<sub>x</sub>), volatile organic compounds (VOC), carbon monoxide (CO), black carbon (BC), organic carbon (OC), and ammonia (NH<sub>3</sub>). Representation of non-CO<sub>2</sub> gases in MESSAGE is described in detail in ref. 63. Here, we primarily focus on CH<sub>4</sub>.

CH<sub>4</sub> emissions are calibrated for the base-year to the RCP inventory<sup>18</sup>. The model represents CH<sub>4</sub> sources by linking appropriate emission coefficients to various activity variables in the model. These include coal, oil and gas extraction and transportation; and energy-related fossil fuel and biomass combustion. We assume gradual technological improvements in regions with high coefficients for these energy-related sources, such as reduced future pipeline leakage in the gas sector in the form of decreased emission coefficients. As explained earlier, in the extraction sector emissions coefficients are different across different natural gas occurrences, and emissions can thus increase when shifting from conventional to unconventional occurrences.

For livestock- and agriculture-related CH<sub>4</sub>, sector-specific drivers are used to project emissions into the future, whereas emissions factors decline for these sources over time, consistent with the projected productivity improvements in livestock management and agricultural production<sup>78</sup>.

For CH<sub>4</sub> emissions from solid waste, we use IPCC country-specific mass-balance methodology<sup>79</sup> to obtain estimates of current emissions. We then examine long-term trends in waste generation rates, recycling, and gas recovery to develop long-term emissions. Based on land availability constraints and current trends in most

developed countries, the rates of recycling and incineration are assumed to increase around the world, thus leading to a lower share of waste on landfills.

MESSAGE considers also the recovery of CH<sub>4</sub> in energy and non-energy sectors. In the energy sector CH<sub>4</sub> may be captured from coal mining (through degasification systems) which is fed into the energy system. In the solid waste sector, the recovered CH<sub>4</sub> from landfills can be used as gas by the industrial sector or converted to electricity for end use. The resulting CH<sub>4</sub> emissions factors of different fossil fuels are shown in Extended Data Table 3.

**Overview of the REMIND model.** REMIND is a global multi-regional model of the energy–economy–climate system spanning the period 2005–2100, with 5-year time steps between 2005 and 2060, and 10-year time steps thereafter. The periods 2005 and 2010 are used for calibration purposes. The scenarios start to differ from 2015 onwards. The world is divided into 11 regions: five individual countries (China, India, Japan, United States of America, and Russia) and six aggregated regions formed by the remaining countries (the European Union, Latin America, sub-Saharan Africa without South Africa, a combined Middle East/North Africa/Central Asia region, other Asia, and the rest of the world).

The macro-economic core of REMIND is an intertemporal general equilibrium model of economic growth with perfect foresight that is solved using optimization methods to compute the market equilibrium with full cooperation between regions. This approach is similar to RICE<sup>80</sup> and MERGE<sup>64</sup>. The macro-economic production function takes as input capital, labour and final energy. The resulting economic output is then available for investments into the macro-economic capital stock as well as for consumption, trade of goods, and financing the energy system. Macro-economic consumption, exogenous population and the pure rate of time preference of 3% per year determine the welfare in each region.

An overview of the REMIND model is shown in Supplementary Fig. 4. The main features are summarized in Supplementary Table 7. The model has been published in the academic literature<sup>81–83</sup> and a full model description is available online<sup>86</sup>.

The REMIND model participated in a number of model comparison studies<sup>30,83–85</sup>. The energy sector and its sub-components have been reviewed in a number of model comparison studies. REMIND performed reasonably compared to the other participating models. REMIND showed particular strength in the sensitivity of price-quantity changes on fossil fuel markets and the international trade of fossil fuels<sup>29,86</sup>. **Modelling energy systems and natural gas in REMIND.** The energy system is hard-linked to the macro-economic core via final energy demand and costs incurred by the energy system<sup>86</sup>. Final energy demand is represented by a production function constant elasticity of substitution (nested CES production function) and includes transport energy, electricity, and various non-electric energy types for stationary end uses. This means that final energy demands are price responsive depending on the substitution elasticities.

The energy sector supplies final energy. The conversion of primary energy into secondary energy carriers as well as the distribution of secondary energy carriers to end-use sectors is represented by capacity stocks of more than 50 technologies in which costs of investment and operation and maintenance are also accounted for. System inertias are represented via the vintage capital structure and adjustment cost for accelerating capacity ramp-up. Therefore, primary energy demands are price elastic and depend on price-elastic final energy demands, all relative prices and system flexibilities in the energy sector. The price-responsive primary energy demands are crucial for the results of this study derived with the REMIND model. The effect of additional gas supplies acts on highly interdependent energy markets and price responsive energy demands are the main trigger for second-order effects due to the gas supply expansion.

The supply side of exhaustible primary energy sources (coal, oil, gas and uranium) assumes cumulative extraction cost functions in each region. In addition to the cumulative extraction cost function the fossil fuel extraction sector distinguishes different grades with an upper limit of supply, specific extraction costs and decline rates. The intertemporal general equilibrium, therefore, reflects producer rents and scarcity rents of the fossil fuel extraction sector. Major subsidies for fossil fuels are also reflected<sup>87</sup>. Natural gas can supply the power sector and supply gases for stationary use reflecting the residential, the service and the industry sector. The fossil fuel sector<sup>16</sup> and the nuclear power sector<sup>82</sup> are fully documented elsewhere. The supply of renewable primary energy comprises renewable energy potentials (biomass, hydro power, wind power, solar energy, and geothermal energy). Renewable energy and storage technologies feature technology learning, reducing investment costs with increasing installed capacity. Furthermore, the integration of fluctuating renewables is subject to integration costs that are implying diminishing returns depending on the market share. Bio-energy supply and land-use emissions are consistent with the land-use model MAGPIE<sup>88</sup>.

International trade is explicitly represented assuming a world market for final goods and primary energy carriers (fossil fuels, uranium and bio-energy). Importers and exporters of primary energy have to pay trading costs, which induce regional price differentials. For the case of natural gas trading costs are substantial and also



energy losses for the transportation are considered. Trading costs are a crucial factor for the diverse development of regional energy systems. A major shift from conventional to abundant gas triggers a change of the relative geo-spatial distribution of gas endowments. In the abundant gas case the endowments are not only larger, but also geographically more evenly distributed, which improves domestic gas supply in many regions that are considered major importers in the conventional gas case.

**Modelling greenhouse gas emissions in REMIND.** Finally the model calculates energy related CO<sub>2</sub> and non-CO<sub>2</sub> greenhouse gas (for example, CH<sub>4</sub> and N<sub>2</sub>O) as well as aerosol emissions via time-dependent emission factors (see Supplementary Table 8). Regarding CH<sub>4</sub> emission factors of extraction activities, they are differentiated for fossil fuels and regions but remain constant over time (see Extended Data Table 3). The marginal abatement cost functions map marginal abatement costs to relative reduction from the baseline level and are linearly scaled with the activity-dependent baseline. These time-dependent marginal abatement cost curves are employed to inform the model about mitigation solutions that prevent CH<sub>4</sub> leakages and make this gas available for supply. Hence, as gas prices increase over time by moving to higher cost deposits, the incentive to invest in mitigation technologies increases and so CH<sub>4</sub> emission factors are effectively reduced. The strength of the effect depends on the endogenous gas price. The greenhouse gas emissions representation in the REMIND model has been fully reported in the peer-reviewed literature<sup>89</sup>.

**Overview of the WITCH model.** WITCH<sup>17</sup> is a dynamic global model that integrates the most important elements of climate change in a unified framework. The economy is modelled through an inter-temporal optimal growth model which captures the long-term economic growth dynamics. A compact representation of the energy sector is fully integrated (hard linked) with the rest of the economy so that energy investments and resources are chosen optimally, together with the other macroeconomic variables. Land-use mitigation options are available through a soft link with a land-use and forestry model (GLOBIOM)<sup>90</sup>. Emission scenarios are processed through a simple climate model calibrated to MAGICC6<sup>22</sup> to compute future climate outcomes. Climate change impacts on the economic output are captured through a damage function, accounting for implicit adaptation decisions. Explicit investment in additional adaptation efforts can reduce the damages associated with temperature change. Feedback loops between economy and climate are thus fully integrated in WITCH to simulate the intertemporal trade-offs between costs of climate change mitigation, adaptation, and residual damages.

WITCH represents the world in a number (currently 13) of representative native regions (or coalitions of regions); as shown in Supplementary Fig. 5, for each it generates optimal mitigation and adaptation strategies for the long term (2005 to 2100), as a result of a maximization process in which the welfare of each region (or coalition of regions) is chosen strategically and simultaneously to other regions. This makes it possible to capture regional free-riding behaviours and strategic interaction induced by the presence of global externalities. In this game-theory set-up, regional strategic actions interrelate through greenhouse gas emissions, dependence on exhaustible natural resources, trade of oil and carbon permits, and technology research and development. The endogenous representation of research-and-development diffusion and innovation processes constitutes a distinguishing feature of WITCH. This approach gives the possibility to explore how research-and-development investments in energy efficiency and carbon-free technologies integrate the currently available mitigation options. The model features multiple externalities, both on the climate and on the innovation side. The technology externality is modelled via international spillovers of knowledge and experience across countries and time<sup>91,92</sup>. This formulation of technical change affects both decarbonization as well as energy savings. Supplementary Table 9 provides an overview of the other key features of WITCH.

**Modelling energy system and natural gas in WITCH.** In WITCH, the energy sector is fully integrated with the rest of the economy. It is distinguished in an electric sector, a transportation sector, and an aggregated non-electric (industry and residential) sectors. The energy sector is described by a production function that aggregates different factors at various levels and with associated elasticities of substitution. All the main energy carriers and technologies are included.

Natural gas is used in the industry and residential sector as well as for generating electricity. Gas power is available with and without carbon capture and storage. WITCH also tracks CH<sub>4</sub> emitted in the non-energy sector. The marginal price of natural gas, along with the other energy carriers, is determined by cumulative global extraction and available resources. Natural gas is traded among the 13 regions, which can buy or sell it from a common pool. Bilateral trade across each region couple is not accounted for. This requires the modeller to vet the trade pattern results carefully when modelling regionally heterogeneous effects. However, this poses little problem in a world of abundant gas availability, where the global gas market is expected to be more integrated and the role of bilateral contracts to be less pronounced.

**Modelling greenhouse gas emissions in WITCH.** The model generates the greenhouse gases reported in Supplementary Table 10, either directly or via exogenous assumptions. Mitigation can happen through technology substitution or storage, direct reduction via marginal abatement cost curves or end of pipe via emission factors. CO<sub>2</sub> emission factors are reported in Extended Data Table 3. Emission trajectories are processed through the MAGICC6 climate model, which calculates the climate outcome.

**Overview of the MAGICC model.** Throughout the analysis we use MAGICC6 for simulating radiative forcing and temperature change. MAGICC is a simple carbon-cycle climate model originally developed by Wigley and Raper<sup>65,93</sup>. The version 6 is updated to emulate the simulations from large-scale climate models and carbon-cycle models as represented in the Coupled Model Intercomparison Project 3 (CMIP3)<sup>40</sup> and Coupled Climate Carbon Cycle Model Intercomparison Project (C4MIP)<sup>94</sup>. See ref. 22 for the documentation of the calibration process.

MAGICC has been traditionally used in the IAMs, and most prominently in the development of the RCPs<sup>95,96</sup>. The RCP scenarios in turn were used in large-scale climate models to simulate the future climate in the IPCC Fifth Assessment Report<sup>19</sup>.

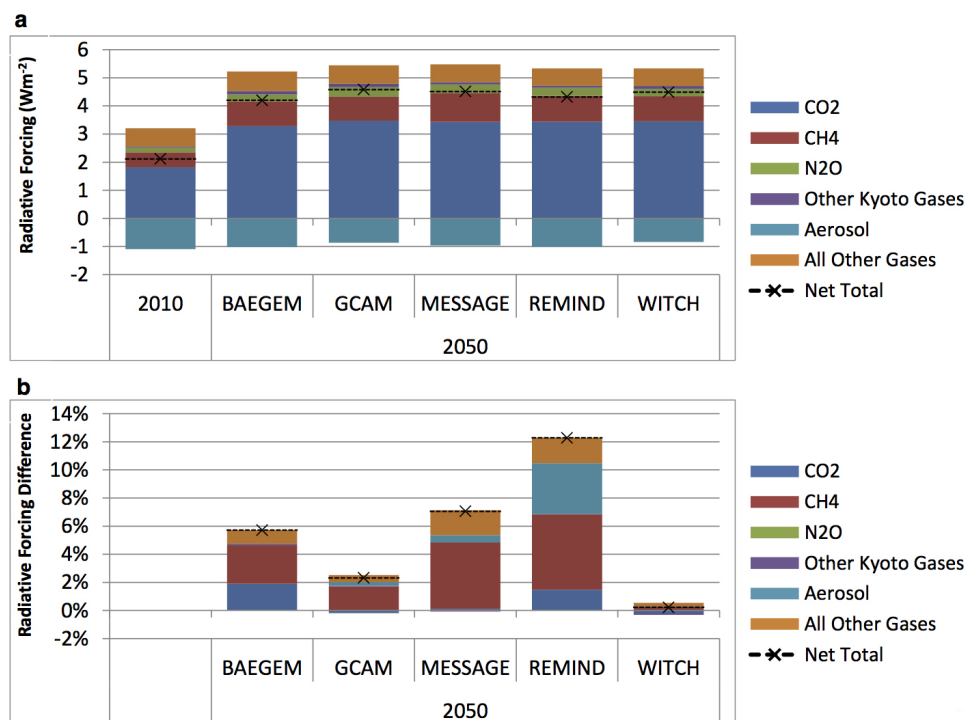
Although a simple climate model like MAGICC is by no means a sufficient substitute for the large-scale climate models, its careful calibration to the large-scale climate models and validation exercises ensure the direction and the magnitude of impact is consistent with the current scientific understanding of the climate system. With its flexible structure and fast runtime, MAGICC can be readily integrated into IAMs. Such a smaller computational burden allows IAMs to simulate more future scenarios without needing a supercomputer to run large-scale climate model.

For this study, we have processed all emission trajectories from the IAMs through MAGICC6 to obtain radiative forcing and temperature change. MAGICC allows emulation of a number of climate models. This study uses the default setting used for RCP analysis. The RCP default setting uses median estimates from the climate model inter-comparison exercises CMIP3<sup>40</sup> and C4MIP<sup>94</sup>. For an emulation of carbon cycle model, the C4MIP Bern-CC model<sup>97</sup> was chosen as it represents the middle of range C4MIP results, and for climate sensitivity 3 °C is used. The full documentation of the RCP default setting is available in ref. 96.

The coverage of greenhouses gases and other forcing agents differs widely across five models. The range of forcing agents covered by each model is available in Supplementary Tables 2, 4, 6, 8, and 10. All models endogenously model CO<sub>2</sub>. All but WITCH endogenously model CH<sub>4</sub> and N<sub>2</sub>O. GCAM, MESSAGE, and REMIND endogenously model aerosols and other short-lived species. For the minor forcing agents that the models do not endogenously simulate, we have used exogenous emissions trajectories from the RCP8.5 scenario<sup>96</sup> because it best approximates our baseline scenario. Forcing the secondary effect of emissions, such as indirect cloud formation or atmospheric chemistry feedback from non-methane hydrocarbons and other reactive gases, is modelled natively in MAGICC<sup>98</sup>.

31. Riahi, K. *et al.* Locked into Copenhagen pledges—implications of short-term emission targets for the cost and feasibility of long-term climate goals. *Technol. Forecast. Soc. Change* (in the press).
32. Nordhaus, W. D. & Boyer, J. G. Requiem for Kyoto: an economic analysis of the Kyoto Protocol. *Energy J.* **20**, 93–130 (1999).
33. Skone, T. J. *Life Cycle Analysis: Natural Gas Combined Cycle (NGCC) Power Plant* (National Energy Technology Laboratory, 2010); <http://www.netl.doe.gov/File%20Library/Research/Energy%20Analysis/Life%20Cycle%20Analysis/NGCC-LCA-Final-Report---Report---9-30-10---Final---Rev-2.pdf>.
34. McCollum, D., Bauer, N., Calvin, K., Kitous, A. & Riahi, K. Fossil resource and energy security dynamics in conventional and carbon-constrained worlds. *Clim. Change* **123**, 413–426 (2014).
35. Messner, S. & Strubegger, M. *User's Guide for MESSAGE III* (International Institute for Applied Systems Analysis, 1995).
36. Luderer, G., Leimbach, M., Bauer, N. & Kriegler, E. Description of the ReMIND-R model. (Potsdam Institute for Climate Impact Research, 2011); [http://www.pik-potsdam.de/research/sustainable-solutions/models/remind/REMIND\\_Description.pdf](http://www.pik-potsdam.de/research/sustainable-solutions/models/remind/REMIND_Description.pdf).
37. Integrated Assessment Modeling Consortium *IPCC AR5 Scenario Database* (IAMC, 2014); <https://secure.iiasa.ac.at/web-apps/ene/AR5DB/>.
38. Reilly, J. M., Edmonds, J. A., Gardner, R. H. & Brenkert, A. L. Uncertainty analysis of the IEA/ORAU CO<sub>2</sub> emissions model. *Energy J.* **8**, 1–29 (1987).
39. Scott, M. J., Sands, R. D., Edmonds, J., Liebetrau, A. M. & Engel, D. W. Uncertainty in integrated assessment models: modeling with MiniCAM 1.0. *Energy Policy* **27**, 855–879 (1999).
40. Meehl, G., Covey, C., McAvaney, B., Latif, M. & Stouffer, R. Overview of the coupled model intercomparison project (CMIP). *Bull. Am. Meteorol. Soc.* **86**, 89–93 (2005).
41. Hertel, T. W. & Hertel, T. W. *Global Trade Analysis: Modeling and Applications* (Cambridge Univ. Press, 1999).
42. Hanoch, G. CRESH production functions. *Econometrica* **39**, 695–712 (1971).
43. Harrison, J., Horridge, J., Jerie, M. & Pearson, K. *GEMPACK Manual* (2012); <http://www.copsmodels.com/gpmanual.htm>.
44. Aguiar, A., McDougall, R. & Narayanan, B. *Global Trade, Assistance and Production: The GTAP 8 Data Base* (Center for Global Trade Analysis, Purdue Univ., 2012).

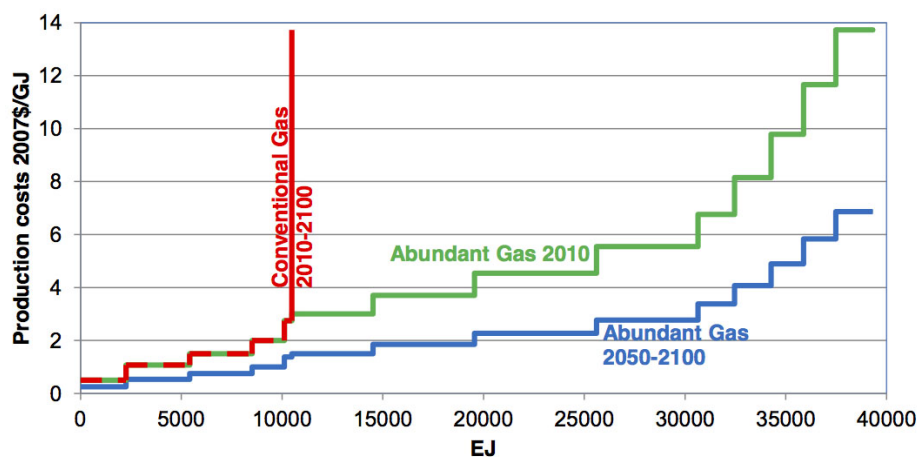
45. International Energy Agency. *CO<sub>2</sub> Emissions from Fuel Combustion 2012* (Organisation for Economic Cooperation and Development, 2012).
46. International Energy Agency. *Emissions of CO<sub>2</sub>, CH<sub>4</sub>, N<sub>2</sub>O, HFCs, PFCs and SF<sub>6</sub>* (Organisation for Economic Cooperation and Development, 2012).
47. United Nations Framework Convention on Climate Change. *2012 Annex I Party GHG Inventory Submissions* (UNFCCC, 2012).
48. Environmental Protection Agency. *Global Anthropogenic Non-CO<sub>2</sub> Greenhouse Gas Emissions: 1990-2030* (US Environmental Protection Agency, 2012).
49. Wigley, T. M. *MAGICC/SCENGEN 5.3: User Manual Version 2* (NCAR, 2008).
50. International Energy Agency. *CO<sub>2</sub> Emissions from Fuel Combustion 1971-2001* (Organisation for Economic Cooperation and Development, 2012).
51. Weyant, J. P., de la Chesnaye, F. C. & Blanford, G. J. Overview of EMF-21: multigas mitigation and climate policy. *Energy J.* **27**, 1-32 (2006).
52. Edmonds, J. & Reilly, J. Global energy production and use to the year 2050. *Energy* **8**, 419-432 (1983).
53. Edmonds, J. & Reilly, J. Global energy and CO<sub>2</sub> to the year 2050. *Energy J.* **4**, 21-47 (1983).
54. Edmonds, J. & Reilly, J. A long-term global energy-economic model of carbon dioxide release from fossil fuel use. *Energy Econ.* **5**, 74-88 (1983).
55. Wise, M. *et al.* Implications of limiting CO<sub>2</sub> concentrations for land use and energy. *Science* **324**, 1183-1186 (2009).
56. Thomson, A. M. *et al.* RCP4.5: a pathway for stabilization of radiative forcing by 2100. *Clim. Change* **109**, 77-94 (2011).
57. Clarke, J. F. & Edmonds, J. A. Modelling energy technologies in a competitive market. *Energy Econ.* **15**, 123-129 (1993).
58. Boden, T. A., Marland, G. & Andres, R. J. *Global, Regional, and National Fossil-Fuel CO<sub>2</sub> Emissions* (Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, US Department of Energy, 2009).
59. International Energy Agency. *Energy Balances of OECD Countries* (Organisation for Economic Cooperation and Development, 2013).
60. International Energy Agency. *Energy Balances of Non-OECD Countries* (Organisation for Economic Cooperation and Development, 2013).
61. Lamarque, J.-F. *et al.* Historical (1850-2000) gridded anthropogenic and biomass burning emissions of reactive gases and aerosols: methodology and application. *Atmos. Chem. Phys.* **10**, 7017-7039 (2010).
62. Messner, S. & Schrattenholzer, L. MESSAGE-MACRO: linking an energy supply model with a macroeconomic module and solving it iteratively. *Energy* **25**, 267-282 (2000).
63. Rao, S. & Riahi, K. The role of non-CO<sub>2</sub> greenhouse gases in climate change mitigation: long-term scenarios for the 21st century. *Energy J.* **27**, 177-200 (2006).
64. Manne, A., Mendelsohn, R. & Richels, R. MERGE: a model for evaluating regional and global effects of GHG reduction policies. *Energy Policy* **23**, 17-34 (1995).
65. Wigley, T. & Raper, S. Implications for climate and sea level of revised IPCC emissions scenarios. *Nature* **357**, 293-300 (1992).
66. Arthur, W. B. Competing technologies, increasing returns, and lock-in by historical events. *Econ. J.* **99**, 116-131 (1989).
67. O'Neill, B. C., Riahi, K. & Keppo, I. Mitigation implications of midcentury targets that preserve long-term climate policy options. *Proc. Natl Acad. Sci. USA* **107**, 1011-1016 (2010).
68. Riahi, K. *et al.* Energy pathways for sustainable development. In *Global Energy Assessment 1203-1306* (Cambridge Univ. Press, 2012).
69. International Energy Agency. *World Energy Outlook 2008* (International Energy Agency, 2008).
70. Riahi, K., Rubin, E. S. & Schrattenholzer, L. Prospects for carbon capture and sequestration technologies assuming their technological learning. *Energy* **29**, 1309-1318 (2004).
71. Ramage, M. & Katzer, J. *Liquid Transportation Fuels from Coal and Biomass: Technological Status, Costs, and Environmental Impacts* (National Academies Press, 2009).
72. Liu, G., Larson, E. D., Williams, R. H., Kreutz, T. G. & Guo, X. Making Fischer-Tropsch fuels and electricity from coal and biomass: performance and cost analysis. *Energy Fuels* **25**, 415-437 (2011).
73. Liu, G., Williams, R. H., Larson, E. D. & Kreutz, T. G. Design/economics of low-carbon power generation from natural gas and biomass with synthetic fuels co-production. *Energy Procedia* **4**, 1989-1996 (2011).
74. Global Energy Assessment GEA Scenario Database (2012); <http://www.iiasa.ac.at/web-apps/ene/geadb/>.
75. Rogner, H.-H. An assessment of world hydrocarbon resources. *Annu. Rev. Energy Environ.* **22**, 217-262 (1997).
76. Patwardhan, A. P., Gomez-Echeverri, L., Johansson, T. B. & Nakićenović, N. *Global Energy Assessment: Toward a Sustainable Future* (Cambridge Univ. Press, 2012).
77. Weber, C. L. & Clavin, C. Life cycle carbon footprint of shale gas: review of evidence and implications. *Environ. Sci. Technol.* **46**, 5688-5695 (2012).
78. Rao, S. & Riahi, K. The role of non-CO<sub>2</sub> greenhouse gases in climate change mitigation: long-term scenarios for the 21st century. *Energy J.* **27**, 177-200 (2006).
79. Intergovernmental Panel on Climate Change. *Guidance and Uncertainty Management in National Greenhouse Gas Inventories 2000* (IPCC Report, 2000).
80. Nordhaus, W. D. & Yang, Z. A regional dynamic general-equilibrium model of alternative climate-change strategies. *Am. Econ. Rev.* **86**, 741-765 (1996).
81. Leimbach, M., Bauer, N., Baumstark, L., Luken, M. & Edenhofer, O. Technological change and international trade-insights from REMIND-R. *Energy J.* **31**, 109-136 (2010).
82. Bauer, N., Brecha, R. J. & Luderer, G. Economics of nuclear power and climate change mitigation policies. *Proc. Natl Acad. Sci. USA* **109**, 16805-16810 (2012).
83. Luderer, G. *et al.* The economics of decarbonizing the energy system—results and insights from the RECIPE model intercomparison. *Clim. Change* **114**, 9-37 (2012).
84. Edenhofer, O. *et al.* The economics of low stabilization: model comparison of mitigation strategies and costs. *Energy J.* **31**, 11-48 (2010).
85. Kriegler, E. *et al.* The role of technology for achieving climate policy objectives: overview of the EMF 27 study on global technology and climate policy strategies. *Clim. Change* **123**, 353-367 (2014).
86. Bauer, N., Edenhofer, O. & Kypreos, S. Linking energy system and macroeconomic growth models. *Comput. Manage. Sci.* **5**, 95-117 (2008).
87. Schwanitz, V. J., Piontek, F., Bertram, C. & Luderer, G. Long-term climate policy implications of phasing out fossil fuel subsidies. *Energy Policy* **67**, 882-894 (2014).
88. Klein, D. *et al.* The value of bioenergy in low stabilization scenarios: an assessment using REMIND-MAGPIE. *Clim. Change* **123**, 705-718 (2013).
89. Strefler, J., Luderer, G., Aboumahboub, T. & Kriegler, E. Economic impacts of alternative greenhouse gas emission metrics: a model-based assessment. *Clim. Change* (in the press).
90. International Institute for Applied Systems Analysis. *Global Biosphere Management Model* (IIASA, 2014); <http://www.iiasa.ac.at/web/home/research/modelsData/GLOBIOM/GLOBIOM.en.html>.
91. Bosetti, V., Carraro, C., Massetti, E. & Tavoni, M. International energy R&D spillovers and the economics of greenhouse gas atmospheric stabilization. *Energy Econ.* **30**, 2912-2929 (2008).
92. Bosetti, V., Carraro, C., Duval, R. & Tavoni, M. What should we expect from innovation? A model-based assessment of the environmental and mitigation cost implications of climate-related R&D. *Energy Econ.* **33**, 1313-1320 (2011).
93. Wigley, T. M. & Raper, S. Thermal expansion of sea water associated with global warming. *Nature* **330**, 127-131 (1987).
94. Friedlingstein, P. *et al.* Climate-carbon cycle feedback analysis: results from the C4MIP model intercomparison. *J. Clim.* **19**, 3337-3353 (2006).
95. Moss, R. H. *et al.* The next generation of scenarios for climate change research and assessment. *Nature* **463**, 747-756 (2010).
96. Meinshausen, M. *et al.* The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Clim. Change* **109**, 213-241 (2011).
97. Joos, F. *et al.* Global warming feedbacks on terrestrial carbon uptake under the Intergovernmental Panel on Climate Change (IPCC) emission scenarios. *Glob. Biogeochem. Cycles* **15**, 891-907 (2001).
98. Wigley, T., Smith, S. J. & Prather, M. Radiative forcing due to reactive gas emissions. *J. Clim.* **15**, 2690-2696 (2002).



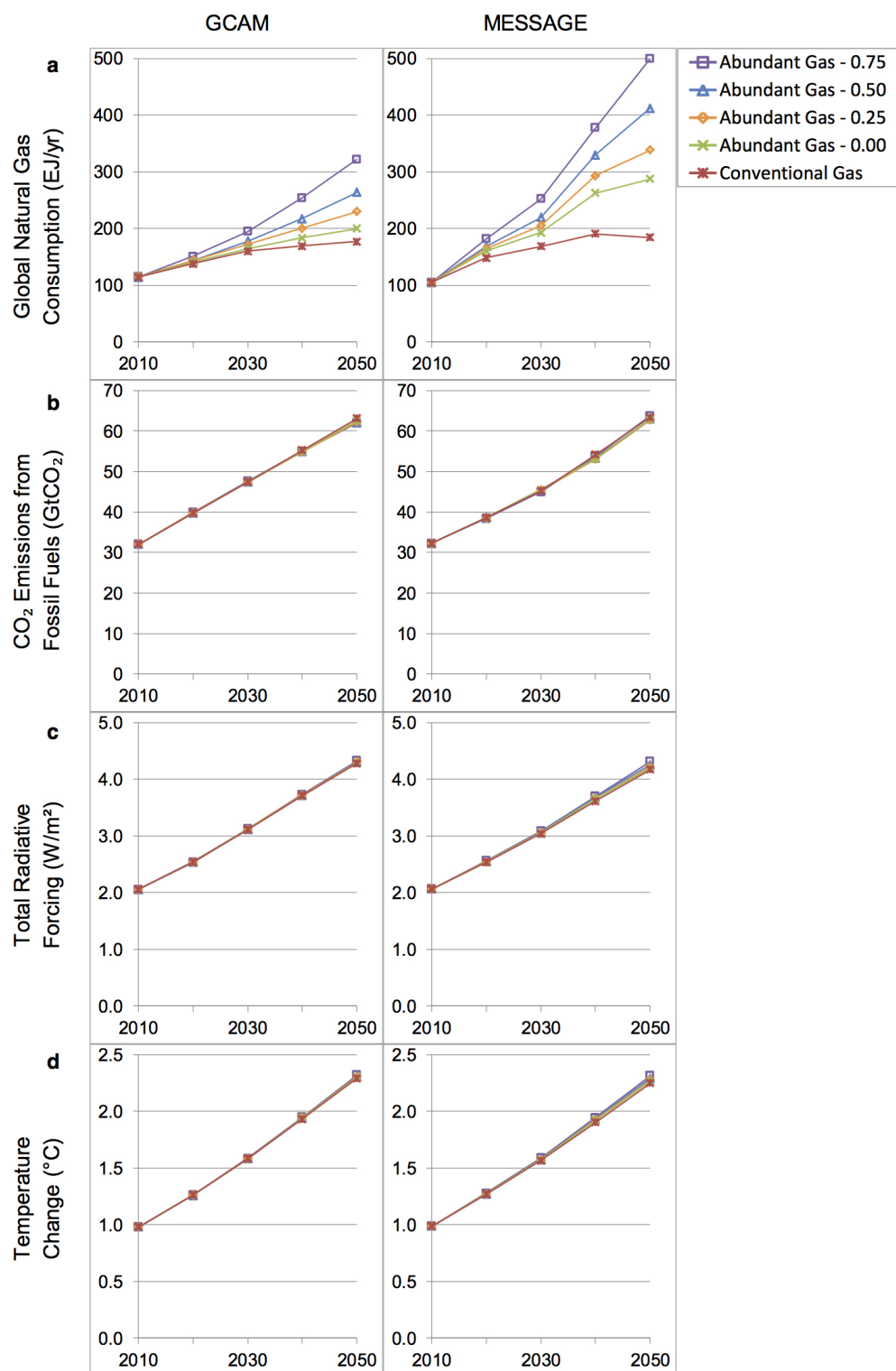
**Extended Data Figure 1 | Radiative forcing composition for high fugitive methane scenarios.** **a**, Year 2010 and year 2050 composition of radiative forcing for the Conventional Gas scenario with high fugitive methane for five models. **b**, Year 2050 relative difference in radiative forcing (the Abundant Gas

scenario minus the Conventional Gas scenario) all with high fugitive methane assumption for the five models. 1% difference in forcing for model average is equivalent to  $0.044 \text{ W m}^{-2}$ .





**Extended Data Figure 2 | Global natural gas supply curves.** The current natural gas supply curves provided by Global Energy Assessment<sup>12</sup>. Future cost reduction assumptions are documented in the Methods.

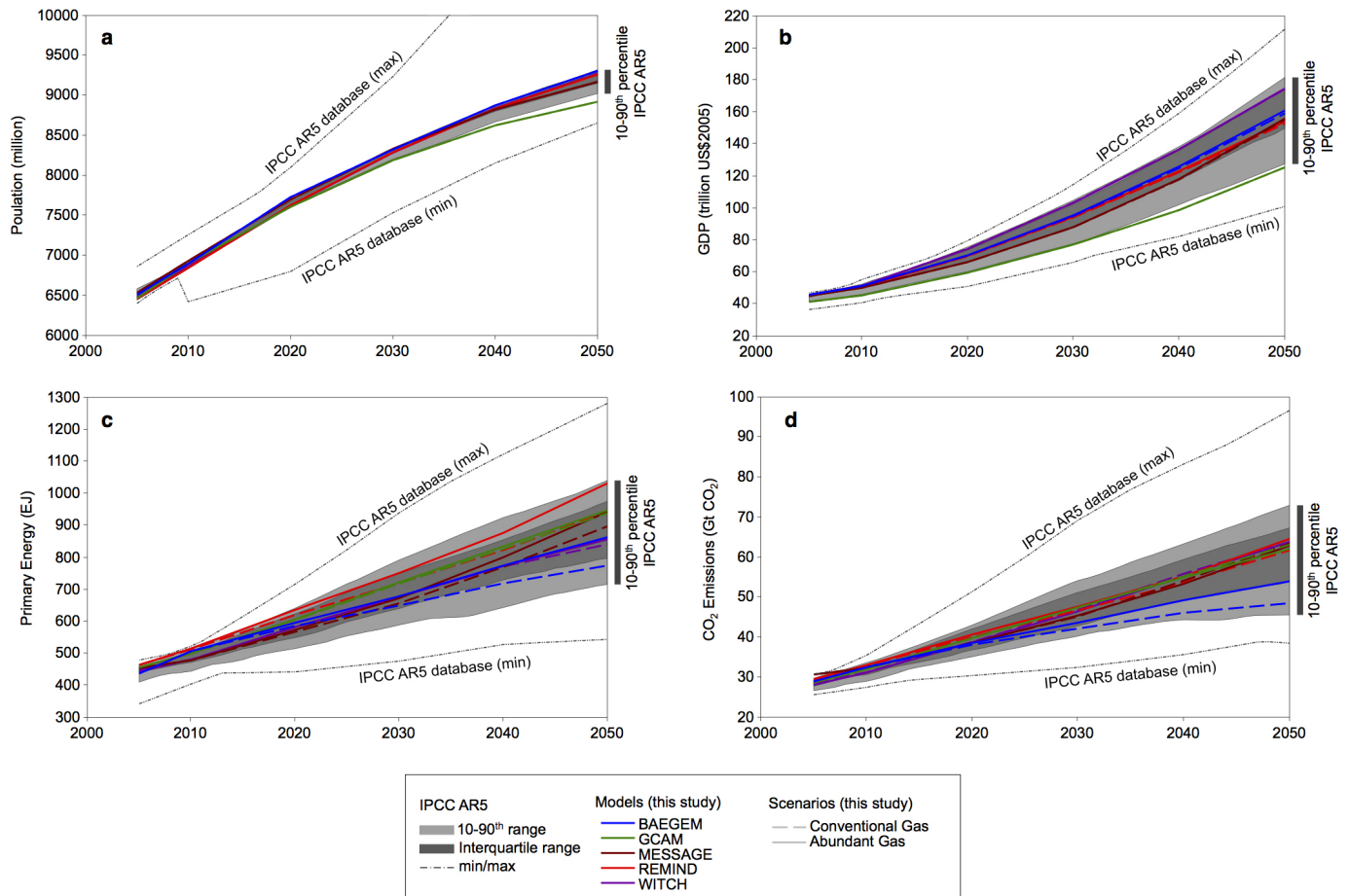


**Extended Data Figure 3 | Natural gas supply curve sensitivity analysis.**

**a**, Global natural gas consumption. **b**, CO<sub>2</sub> emissions from fossil fuels.

**c**, Total radiative forcing. **d**, Global mean surface temperature change (from

pre-industrial average 1750–1849). Conventional Gas and Abundant Gas denote the quantity of natural gas supply. The decimal numbers denote the fraction of cost reduction over 2010–2050.



**Extended Data Figure 4 | Uncertainty ranges in principal components of model projections.** **a**, Global population. **b**, Global GDP. **c**, Total primary energy consumption. **d**, Fossil fuel and industrial CO<sub>2</sub> emissions. Coloured

lines are model reported values from this study. Shaded areas are ranges of projections found in the literature obtained from the IPCC AR5 database<sup>37</sup>.



**Extended Data Table 1 | Cost reduction in low-carbon energy technologies over 2010–2050 in the Abundant Gas scenario**

	BAEGEM	GCAM	MESSAGE	REMIND	WITCH	units
Solar Photovoltaics	36%	63%	58%	21%	53%	%
Wind Turbine	37%	21%	41%	13%	40%	%
Nuclear Powerplant	29%	15%	17%	0%	7%	%

PV, photovoltaics.

**Extended Data Table 2 | CO<sub>2</sub> emissions in 2050 from fossil fuels and industry with standard energy market assumptions and with the coal-substitution-only assumption**

		BAEGEM	GCAM	MESSAGE	REMIND	WITCH	units
Standard	Conventional Gas	48.5	63.1	63.4	61.7	63.8	GtCO <sub>2</sub>
	Abundant Gas	54.0	62.1	62.9	64.6	63.9	GtCO <sub>2</sub>
	Difference	5.5	-1.1	-0.5	2.9	0.0	GtCO <sub>2</sub>
Coal Substitution Only	Conventional Gas	48.5	63.1	63.4	61.7	63.8	GtCO <sub>2</sub>
	Abundant Gas	51.9	60.4	59.7	60.4	63.8	GtCO <sub>2</sub>
	Difference	3.3	-2.8	-3.8	-1.3	-0.0	GtCO <sub>2</sub>

**Extended Data Table 3 | 2050 emission factors for fossil fuels in each model**

		BAEGEM	GCAM	MESSAGE	REMIND	WITCH	units
CO <sub>2</sub>	Coal	101	100	95-101	96	90	kgCO <sub>2</sub> /GJ
	Oil	79	72	73	68	70	kgCO <sub>2</sub> /GJ
	Gas	59	52	56	56	55	kgCO <sub>2</sub> /GJ
CH <sub>4</sub>	Coal	0.21	0.14	0.39	0.12	N/A	kgCH <sub>4</sub> /GJ
	Oil	0.11	0.06	0.06	0.06	N/A	kgCH <sub>4</sub> /GJ
	Gas	0.32	0.35	0.31	0.52	N/A	kgCH <sub>4</sub> /GJ

CO<sub>2</sub> emission factors specify the average carbon content of the fuel. CH<sub>4</sub> emission factors specify average fugitive methane emissions associated with production and transportation of each fossil fuel reported for the Abundant Gas scenario.



**Extended Data Table 4 | 2050 anthropogenic radiative forcing with standard fugitive methane emission assumptions and with high fugitive methane emission assumptions**

		BAEGEM	GCAM	MESSAGE	REMIND	WITCH	units
Standard	Conventional Gas	3.97	4.46	4.25	4.16	4.38	W m <sup>-2</sup>
	Abundant Gas	4.07	4.49	4.37	4.46	4.37	W m <sup>-2</sup>
	Difference	0.10	0.02	0.12	0.31	-0.01	W m <sup>-2</sup>
High Fugitive Methane	Conventional Gas	4.20	4.58	4.51	4.32	4.49	W m <sup>-2</sup>
	Abundant Gas	4.44	4.69	4.83	4.85	4.50	W m <sup>-2</sup>
	Difference	0.24	0.11	0.32	0.53	0.01	W m <sup>-2</sup>

# Producing more grain with lower environmental costs

Xinping Chen<sup>1\*</sup>, Zhenling Cui<sup>1\*</sup>, Mingsheng Fan<sup>1</sup>, Peter Vitousek<sup>2</sup>, Ming Zhao<sup>3</sup>, Wenqi Ma<sup>4</sup>, Zhenlin Wang<sup>5</sup>, Weijian Zhang<sup>3</sup>, Xiaoyuan Yan<sup>6</sup>, Jianchang Yang<sup>7</sup>, Xiping Deng<sup>8</sup>, Qiang Gao<sup>9</sup>, Qiang Zhang<sup>10</sup>, Shiwei Guo<sup>11</sup>, Jun Ren<sup>12</sup>, Shiqing Li<sup>8</sup>, Youliang Ye<sup>13</sup>, Zhaohui Wang<sup>14</sup>, Jianliang Huang<sup>15</sup>, Qiyuan Tang<sup>16</sup>, Yixiang Sun<sup>17</sup>, Xianlong Peng<sup>18</sup>, Jiawang Zhang<sup>5</sup>, Mingrong He<sup>5</sup>, Yunji Zhu<sup>13</sup>, Jiquan Xue<sup>14</sup>, Guiliang Wang<sup>1</sup>, Liang Wu<sup>1</sup>, Ning An<sup>1</sup>, Liangquan Wu<sup>1</sup>, Lin Ma<sup>1</sup>, Weifeng Zhang<sup>1</sup> & Fusuo Zhang<sup>1</sup>

**Agriculture faces great challenges to ensure global food security by increasing yields while reducing environmental costs<sup>1,2</sup>. Here we address this challenge by conducting a total of 153 site-year field experiments covering the main agro-ecological areas for rice, wheat and maize production in China. A set of integrated soil–crop system management practices based on a modern understanding of crop ecophysiology and soil biogeochemistry increases average yields for rice, wheat and maize from 7.2 million grams per hectare ( $\text{Mg ha}^{-1}$ ), 7.2  $\text{Mg ha}^{-1}$  and 10.5  $\text{Mg ha}^{-1}$  to 8.5  $\text{Mg ha}^{-1}$ , 8.9  $\text{Mg ha}^{-1}$  and 14.2  $\text{Mg ha}^{-1}$ , respectively, without any increase in nitrogen fertilizer. Model simulation and life-cycle assessment<sup>3</sup> show that reactive nitrogen losses and greenhouse gas emissions are reduced substantially by integrated soil–crop system management. If farmers in China could achieve average grain yields equivalent to 80% of this treatment by 2030, over the same planting area as in 2012, total production of rice, wheat and maize in China would be more than enough to meet the demand for direct human consumption and a substantially increased demand for animal feed, while decreasing the environmental costs of intensive agriculture.**

Global agriculture is facing unprecedented challenges and risks. Rates of yield growth have slowed since the 1980s (ref. 4), and even stagnated in many areas<sup>5–7</sup>. Meanwhile, agriculture incurs substantial environmental costs, including emissions of greenhouse gases<sup>8</sup>, loss of biodiversity<sup>9</sup>, and degradation of land and freshwater<sup>10,11</sup>. These challenges may grow in the future, because global food demand is likely to double by 2050 (reflecting both population growth and increased consumption of animal protein) against a backdrop of a changing climate and growing competition for land, water, labour and energy<sup>2</sup>. The human and environmental costs of expanding agricultural lands are such that most of the necessary production gains must be achieved on existing farmland<sup>1</sup>. Can the necessary increase in yields be accomplished? If so, can the environmental costs of intensive agriculture be mitigated?

We addressed these questions through quantitative field experiments under large-scale agro-ecological conditions. Our experiments included the three main staple crops (rice, wheat and maize), which together account for most global cereal production<sup>12,13</sup>, in the main agro-ecological areas of China. We focus on China in part because the yields of these crops already are relatively high there, thanks to ‘green revolution’ technologies, and in part because China must address the joint challenges of production and environmental degradation expeditiously<sup>14</sup>.

From 2009 to 2012, we conducted a total of 153 site-year field experiments (Extended Data Fig. 1). In each experiment four treatments were employed: (1) current practice (the farmers’ practice in the region but

conducted in experimental plots); (2) improved practice (which modified current practice to offset the major limitations to crop growth); (3) high-yielding (which maximized yields without regard to costs); and (4) integrated soil–crop system management (ISSM, which used advanced crop and nutrient management). ISSM redesigned the whole production system based on the local environment, drawing upon appropriate crop varieties, sowing dates, densities and advanced nutrient management. The ISSM concept had been developed for maize systems<sup>15</sup>; we applied it to a broad range of field situations for wheat and rice in addition to maize. The challenge of increasing yield while reducing environmental costs is greater for tiller crops such as rice and wheat because they change in population structure within crop growing seasons (Supplementary Discussion). In addition to our experiments, we determined yields and nitrogen use in 18,938 farmers’ fields in the main cereal production areas of China (Extended Data Table 1).

Our highest yields were achieved in high-yielding treatments, with 8.8, 9.2 and 14.4  $\text{Mg ha}^{-1}$  for rice, wheat and maize, respectively (Table 1). These yields are comparable to yield potentials in the areas with the most favourable conditions and intensive agronomic management globally: rice in California (USA) ( $\sim 9 \text{ Mg ha}^{-1}$ ) (ref. 5), wheat in Germany (9.5  $\text{Mg ha}^{-1}$ ), and rainfed and irrigated maize in the USA (13.2 and 15.1  $\text{Mg ha}^{-1}$ ) (ref. 16). The ISSM treatment achieved 97–99% of the yields in the high-yielding treatments, and the improved practice treatments achieved 88–92% of high-yielding yields; all increased yields (Table 1) and nitrogen uptakes were significantly greater than the current practice treatment (Extended Data Table 2).

Nitrogen fertilizer application rates were greatest in the high-yielding treatment, and decreased in the order current practice then ISSM then improved practice (Table 1). High nitrogen surplus (nitrogen fertilizer applied in excess of uptake by crops) and low nitrogen use efficiency (PFP<sub>N</sub>, nitrogen partial factor productivity, in kilograms of grain per kilogram of nitrogen applied) occurred in high-yielding (owing to high nitrogen application) and current practice (owing to low grain yield) treatments, indicating the inefficiency and environmental damage associated with both conventional practices and with attempts to increase yields simply by increasing inputs<sup>17</sup>. Compared with current practice, the nitrogen rates in improved practice and ISSM decreased slightly even as yields increased substantially.

In the improved practice and ISSM treatments, nitrogen surplus was around zero with only a small range from  $-9$  to 16  $\text{kg N ha}^{-1}$ , and PFP<sub>N</sub> reached 54–57, 41–44 and 56–59  $\text{kg N kg}^{-1}$  for rice, wheat and maize, respectively (Table 1). These nitrogen use efficiencies are comparable to those of most ‘ecologically intensive’ systems worldwide<sup>18</sup>.

<sup>1</sup>College of Resources & Environmental Sciences, China Agricultural University, Beijing 100193, China. <sup>2</sup>Department of Biology, Stanford University, Stanford, California 94305, USA. <sup>3</sup>Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing 100081, China. <sup>4</sup>College of Resources & Environmental Sciences, Agricultural University of Hebei, Baoding 071001, China. <sup>5</sup>College of Agronomy, Shandong Agricultural University, Tai'an 271000, China. <sup>6</sup>Institute of Soil Science, Chinese Academy of Sciences, Nanjing 210008, China. <sup>7</sup>Key Laboratory of Crop Genetics and Physiology of Jiangsu Province, Yangzhou University, Yangzhou 225009, China. <sup>8</sup>State Key Laboratory of Soil Erosion and Dryland Farming on the Loess Plateau, Northwest Agriculture and Forestry University, Yangling 712100, China. <sup>9</sup>College of Resources & Environmental Sciences, Jilin Agricultural University, Changchun 130118, China. <sup>10</sup>Institute of Agricultural Environment and Resource, Shanxi Academy of Agricultural Sciences, Taiyuan 030031, China. <sup>11</sup>College of Resources & Environmental Sciences, Nanjing Agricultural University, Nanjing 210095, China. <sup>12</sup>Research Center of Agricultural Environment & Resources, Jilin Academy of Agricultural Sciences, Changchun 130033, China. <sup>13</sup>College of Resources & Environmental Sciences, Henan Agricultural University, Zhengzhou 450000, China. <sup>14</sup>Northwest Agriculture and Forestry University, Yangling 712100, China. <sup>15</sup>College of Plant Science & Technology, Huazhong Agricultural University, Wuhan 430070, China. <sup>16</sup>Crop Physiology, Ecology & Production Center, Hunan Agricultural University, Changsha 410128, China. <sup>17</sup>Soil & Fertilizer Research Institute, Anhui Academy of Agricultural Sciences, Hefei 230031, China. <sup>18</sup>College of Resources & Environmental Sciences, Northeast Agricultural University, Harbin 150030, China.

\*These authors contributed equally to this work.

**Table 1 | Grain yield, nitrogen application rate, PFP<sub>N</sub> and nitrogen surplus for rice (*n* = 57), wheat (*n* = 40) and maize systems (*n* = 56) for the four management treatments in field experiments compared with farmers' practice from a total of 18,938 farmers**

Crops	Treatment	Yield (Mg ha <sup>-1</sup> )	N rate (kg N ha <sup>-1</sup> )	PFP <sub>N</sub> (kg kg <sup>-1</sup> )	N surplus (kg N ha <sup>-1</sup> )
Rice	Current practice	7.2 ± 1.1‡	181*	41§	58*
	Improved practice	8.1 ± 1.1†	146‡	57*	7‡
	High-yielding system	8.8 ± 1.2*	192*	47‡	38†
	ISSM	8.5 ± 1.2*†	162†	54†	16‡
	Farmers' practice ( <i>n</i> = 6,592)	7.0 ± 1.5	209	41	82
Wheat	Current practice	7.2 ± 1.4‡	257†	28‡	74*
	Improved practice	8.3 ± 1.7†	192§	44*	-9†
	High-yielding system	9.2 ± 1.9*	283*	33†	50*
	ISSM	8.9 ± 1.7*†	220‡	41*	2†
	Farmers' practice ( <i>n</i> = 6,940)	5.7 ± 1.3	210	33	74
Maize	Current practice	10.5 ± 1.6‡	266†	40†	72†
	Improved practice	12.6 ± 2.2†	214‡	59*	-8‡
	High-yielding system	14.4 ± 2.4*	402*	37†	140*
	ISSM	14.2 ± 2.6*	256†	56*	8‡
	Farmers' practice ( <i>n</i> = 5,406)	7.6 ± 1.5	220	43	72

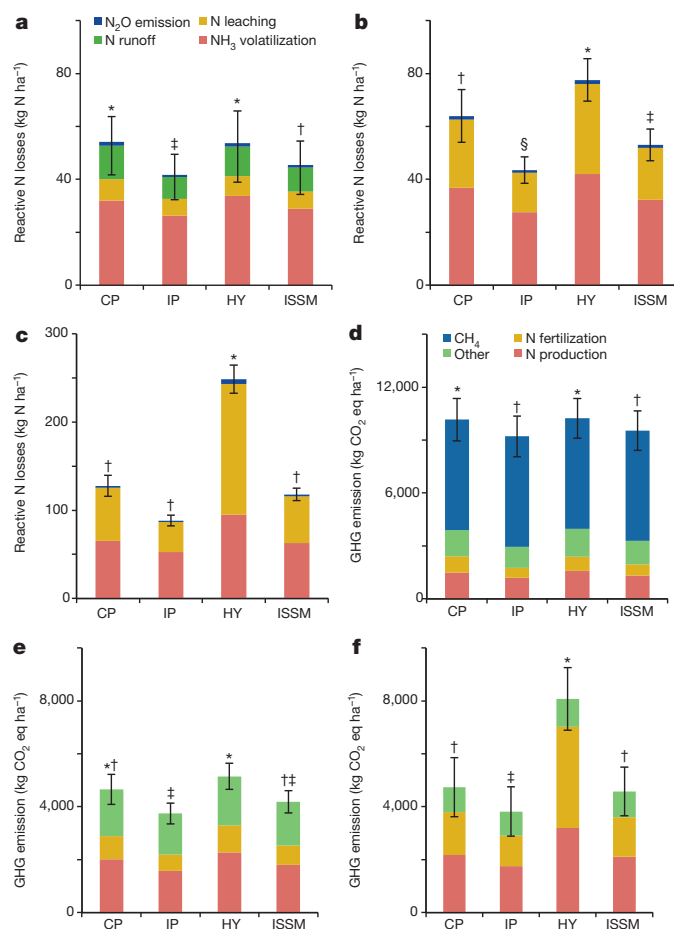
Means ± s.d. for yield. Least significant difference testing was performed among the four experimental treatments for each crop, the same footnote symbol(s) within each column are not significantly different at *P* < 0.05.

Reactive nitrogen losses and greenhouse gas (GHG) emissions from agriculture contribute substantially to atmospheric and water pollution in China and elsewhere. Using established empirical models<sup>19,20</sup> (Extended Data Figs 2–4 and Supplementary Discussion) and life-cycle assessment methods<sup>3,19</sup>, we evaluated total reactive nitrogen losses and GHG emissions per unit area (expressed as kilograms of nitrogen per hectare or kilograms of carbon dioxide equivalents per hectare), and their intensity per unit grain yield (expressed as kilograms of nitrogen losses or carbon dioxide equivalents per million grams). Total reactive nitrogen losses and GHG emissions both for improved practice and ISSM treatments decreased compared with current practice, while high-yielding significantly increased total reactive nitrogen losses and GHG emissions (except GHG emissions in rice systems) (Fig. 1).

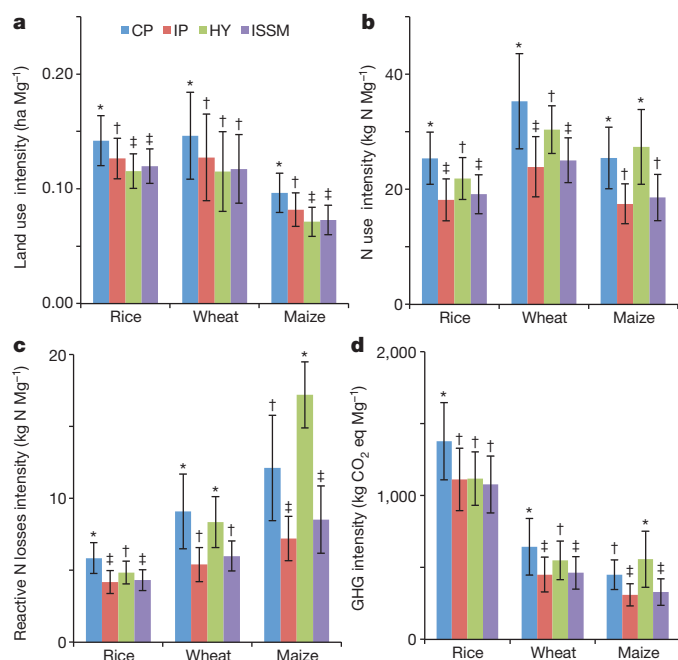
Reactive nitrogen losses in maize systems were higher than those in wheat and rice systems (Fig. 1), mainly because of high nitrate leaching and ammonia volatilization in maize's summer growing season. The total GHG emissions from rice were highest because of high methane (CH<sub>4</sub>) emissions (Fig. 1). Nitrogen fertilizer production, transportation and application contributed substantially to the difference in total GHG emissions among treatments, especially for wheat and maize.

These large gains in grain yield for maize, wheat and rice demonstrate a substantial potential to meet food demand on existing farmland. While previous calculations based on historic yield trends suggested strong constraints—for example, suggesting that Chinese rice yields reached a plateau of ~6.4 Mg ha<sup>-1</sup> by the mid-1990s (ref. 5)—our results demonstrate that this plateau does not represent a biophysical yield limitation (yield ceiling). We suggest that socio-economic factors—particularly extremely small farm sizes and urbanization leading to an increase in the proportion of part-time farmers—contribute to the observed yield plateau (Extended Data Fig. 5). These socio-economic factors could diminish with economic development and changes in land tenure or management arrangements (Supplementary Discussion). More generally, while previous calculations based on historic yield trends suggested that crop yields have reached a plateau in much of the world<sup>5–7</sup>, our large-scale experimental results demonstrate that this suggestion requires careful testing.

Equally importantly, our experiments demonstrate that substantially increased yields can be produced with lower inputs of nitrogen fertilizer, and so lower human and environmental costs (Fig. 2). Our survey of farmers also shows that it is possible to achieve high yields in practice, because we found that about 20% and 5% of rice and wheat farmers, respectively, report yields already close to ISSM yields without using excessive fertilizer (Extended Data Table 3). Even so, there is room for further improvement: GHG emission intensities in ISSM (our best treatment) are still higher than published results in other intensive agricultural regions, such as maize in the USA (231 kg CO<sub>2</sub> eq Mg<sup>-1</sup> of grain, and 13.2 Mg grain ha<sup>-1</sup>)<sup>20</sup>, mainly because of coal-based nitrogen fertilizer



**Figure 1 | Reactive nitrogen losses and GHG emissions for four management treatments, based on empirical models of losses and life-cycle assessment.** a–c, Reactive nitrogen losses (a, rice; b, wheat; c, maize) include N<sub>2</sub>O emission, nitrogen leaching, NH<sub>3</sub> volatilization and nitrogen runoff. d–f, GHG emissions (d, rice; e, wheat; f, maize) include those from nitrogen fertilizer application, nitrogen fertilizer production and transportation, other sources (phosphorus and potassium fertilizer; crop management) and CH<sub>4</sub> emission (in rice). CP, current practice; IP, improved practice; HY, high-yielding system. Means followed by the same footnote symbol(s) for each crop are not significantly different at *P* < 0.05.

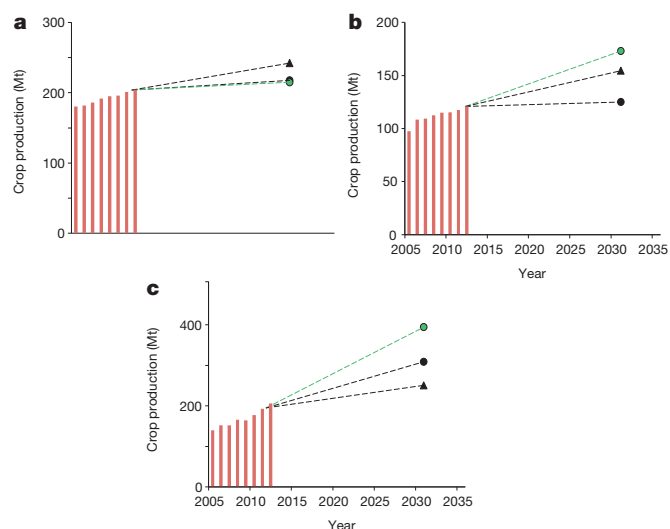


**Figure 2 | Substantially increased yields can be produced with lower inputs of nitrogen fertilizer, and so lower human and environmental costs.** The intensity of land use (a), nitrogen use (b), reactive nitrogen losses (c) and GHG emissions (d) needed to produce 1 Mg of grain, for three crops and four management treatments. Means followed by the same footnote symbol(s) for each crop are not significantly different at  $P < 0.05$ .

production in China<sup>21</sup>. Improved fertilizer production technology<sup>21</sup> and innovative fertilizer products<sup>22</sup> could play further roles in mitigating GHG emissions.

Current yields and cropping areas across China combine to produce 204, 121 and 206 Mt of rice, wheat and maize annually<sup>23</sup>, with 74% of maize fed to livestock (with 5 Mt of imported maize, and 58 Mt of imported soybean). With population and economic growth, demand for grain in China is expected to reach 218, 125 and 315 Mt of rice, wheat and maize by 2030, by which time China's population is expected to have stabilized. If farmers could achieve grain yields of 80% of the yield level in our ISSM treatment by 2030, using the same planting area as in 2012, total production of rice, wheat and maize would reach 216, 174 and 397 Mt; this is enough to meet the demand for direct human consumption and domestically produced animal feed. Such yields would even suffice to offset imports of animal feed (Fig. 3), while reducing nitrogen use, reactive nitrogen losses and GHG emissions by 21%, 30% and 11% respectively, compared with current levels (scenario 2 in Extended Data Table 4). Further, if we simply reach the projected demand in 2030 with 80% of ISSM yields, then reactive nitrogen losses and GHG emission could be reduced by 48% and 26%, and the land and nitrogen fertilizer used for these three crops could also be reduced by 22% and 33% (scenario 3 in Extended Data Table 4). This change could contribute to the production of other crops and to the protection of natural ecosystems. Also, a relative shift to maize will reduce agricultural demand for water in China<sup>24</sup>. However, if larger quantities of more sustainably produced grain are allocated to an inefficient animal production system, overall benefits will be reduced substantially<sup>25</sup>. Increasing the efficiency and mitigating the environmental/human costs of livestock production systems in China deserves more attention.

The gains in yield and environmental quality that can be achieved through an integrated agronomic approach are striking—especially given that yields in China are already higher than those in most developing countries. The ISSM approach is agronomically robust and relatively easy and inexpensive to adopt (Supplementary Discussion), although the management practices employed for ISSM vary across different crops



**Figure 3 | The projected demand of grain production for 2030 in China.** a, Rice; b, wheat; c, maize. Red bars, crop production from 2005 to 2012. Black circles, projected demand in 2030. Black triangles, increasing grain yield by the trend observed from 2005 to 2012, keeping planting area the same as in 2012. Green circles, grain yields that reach 80% of the level observed in our ISSM treatment, over the same planting area as in 2012. Note differences in scale for the different crops.

and different regions. We believe that this approach can be applied elsewhere—and that it should be possible to meet global food demand with more sustainable intensive agriculture on existing cropland, thereby sustaining other natural resources by avoiding the conversion of forest, grassland and marginal lands to agriculture and supporting other ecosystem services such as wetland preservation, wildlife conservation, carbon sequestration, etc. These benefits are achievable if we invest in agronomic research that incorporates an ecosystem perspective, if the effort is pursued across disciplinary and institutional boundaries, and if we provide the technologies, arrangements and incentives that make it viable for farmers to adapt and adopt more knowledge-intensive forms of agriculture.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 20 May; accepted 25 June 2014.

Published online 3 September 2014.

- Foley, J. A. *et al.* Solutions for a cultivated planet. *Nature* **478**, 337–342 (2011).
- Tilman, D., Balzer, C., Hill, J. & Belfort, B. L. Global food demand and the sustainable intensification of agriculture. *Proc. Natl Acad. Sci. USA* **108**, 20260–20264 (2011).
- Audley, E. *et al.* *Harmonisation of Environmental Life Cycle Assessment for Agriculture*. Final Report. Concerted Action MR3-CT94-2028 (Silsoe Research Institute, 1997).
- Alston, J. M., Beddow, J. M. & Pardey, P. G. Agricultural research, productivity, and food prices in the long run. *Science* **325**, 1209–1210 (2009).
- Grassini, P., Eskridge, K. M. & Cassman, K. G. Distinguishing between yield advances and yield plateaus in historical crop production trends. *Nat. Commun.* **4**, 2918–2928 (2013).
- Ray, D. K., Mueller, N. D., West, P. C. & Foley, J. A. Yield trends are insufficient to double global crop production by 2050. *PLoS ONE* **8**, e66428 (2013).
- Ray, D. K. *et al.* Recent patterns of crop yield growth and stagnation. *Nat. Commun.* **3**, 1293–1299 (2012).
- Davidson, E. A. The contribution of manure and fertilizer nitrogen to atmospheric nitrous oxide since 1860. *Nat. Geosci.* **2**, 659–662 (2009).
- Christopher, M. C. & Tilman, D. Loss of plant species after chronic low-level nitrogen deposition to prairie grasslands. *Nature* **451**, 712–715 (2008).
- Diaz, R. J. & Rosenberg, R. Spreading dead zones and consequences for marine ecosystems. *Science* **321**, 926–929 (2008).
- Guo, J. H. *et al.* Significant acidification in major Chinese croplands. *Science* **327**, 1008–1010 (2010).
- Statistics Division of the Food and Agriculture Organization of the United Nations. Agriculture database. FAOSTAT <http://faostat3.fao.org/faostat-gateway/go/to/home/E> (29 July 2014).
- Cassman, K. G., Dobermann, A. & Walters, D. T. Agroecosystems, nitrogen-use efficiency, and nitrogen management. *Ambio* **31**, 132–140 (2002).



14. Zhang, F. S., Chen, X. P. & Vitousek, P. Chinese agriculture: an experiment for the world. *Nature* **497**, 33–35 (2013).
15. Chen, X. P. *et al.* Integrated soil–crop system management for food security. *Proc. Natl Acad. Sci. USA* **108**, 6399–6404 (2011).
16. Wart, J., Grassini, P. & Cassman, K. G. Impact of derived global weather data on simulated crop yields. *Glob. Change Biol.* **19**, 3822–3834 (2013).
17. Tilman, D. *et al.* Forecasting agriculturally driven global environmental change. *Science* **292**, 281–284 (2001).
18. Dobermann, A. in *IFA Int. Workshop on Enhanced-Efficiency Fertilizers*, 28–30 June 2005, Frankfurt, Germany (International Fertilizer Industry Association, 2005).
19. Cui, Z. L. *et al.* Closing the yield gap could reduce projected greenhouse gas emissions: a case study of maize production in China. *Glob. Change Biol.* **19**, 2467–2477 (2013).
20. Grassini, P. & Cassman, K. G. High-yield maize with large net energy yield and small global warming intensity. *Proc. Natl Acad. Sci. USA* **109**, 1074–1079 (2012).
21. Zhang, W. F. *et al.* New technologies reduce greenhouse gas emissions from nitrogenous fertilizer in China. *Proc. Natl Acad. Sci. USA* **110**, 8375–8380 (2013).
22. Chien, S. H., Prochnow, L. I. & Cantarella, H. Recent developments of fertilizer production and use to improve nutrient efficiency and minimize environmental impacts. *Adv. Agron.* **102**, 267–322 (2009).
23. National Bureau of Statistics of China. *China Statistical Yearbook* (China Statistics, 2013).
24. Zwart, S. J. & Bastiaanssen, G. M. Review of measured crop water productivity values for irrigated wheat, rice, cotton and maize. *Agric. Water Manage.* **69**, 115–133 (2004).
25. Ma, L. *et al.* Environmental assessment of management options for nutrient flows in the food chain in China. *Environ. Sci. Technol.* **47**, 7260–7268 (2013).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank P. A. Matson, G. P. Robertson, I. Ortiz-Monasterio and G. Maltais-Landry for their comments on an earlier version of the manuscript or assistance during the manuscript revision, and we thank C. L. Kou, D. S. Tan, Z. M. Wang, Z. A. Lin, X. Y. Zhang, J. L. Gao and Y. Zhu for joining field experiments. We also acknowledge all those who provided local assistance or technical help to the Integrated Nutrient Management Network in China. This work was financially supported by the Chinese National Basic Research Program (2009CB118600), the Innovative Group Grant from the NSFC (31121062) and the Special Fund for Agro-scientific Research in the Public Interest (201103003).

**Author Contributions** X.C. and F.Z. designed the research. Z.C., Z.W., M.Z., W.M., W.Z., X.Y., J.Y., X.D., Q.G., Q.Z., S.G., J.R., S.L., Y.Y., Z.W., J.H., Q.T., Y.S., X.P., J.Z., M.H., Y.Z. and J.X. conducted field experiments. Z.C., M.F., G.W., L.W., N.A., L.W., L.M. and W.Z. collected the data sets and analysed the data. X.C., Z.C. and P.V. wrote the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to F.Z. ([zhangfs@cau.edu.cn](mailto:zhangfs@cau.edu.cn)).

## METHODS

**Field experiments.** A total 153 site-years of field experiments were conducted from 2009 to 2012 within the main agro-ecological areas for rice ( $n = 57$ ), wheat ( $n = 40$ ) and maize ( $n = 56$ ) production in China (Extended Data Fig. 1). Four treatments were designed and compared: (1) current practice, which followed farmers' practices in the region but was conducted in experimental plots; (2) improved practice, which was based on improving farmers' practices beginning with an analysis of limiting factors, followed by implementing key new technologies, mostly through using root zone nutrient management<sup>26</sup> to improve nutrient use efficiency, together with known agronomic management practices (that is, increasing planting density) to increase yield; (3) high-yielding, designed to test yield potential, where crop yields were maximized through inputs so that they made full use of solar radiation and the period with favourable temperatures, without considering the costs of various inputs; (4) ISSM, which redesigned cropping systems using advanced crop and nutrient management to bring yields closer to their biophysical potential, while optimizing various resource inputs (that is, nutrient and water) and minimizing environmental costs, based on an understanding of crop ecophysiology (for example, crop canopy, solar radiation use and dry matter accumulation), physiological nutrient demands by high-yielding crop and the biogeochemical processes relating to nutrient availability and loss<sup>15</sup>.

A randomized complete block design with four replications was used for each experiment. At maturity, grain yield and above-ground biomass were sampled and measured in each plot, with 6 m<sup>2</sup> for wheat and rice, and 10 m<sup>2</sup> for maize. Their nitrogen concentrations were determined using the Kjeldahl procedure. Fertilizer and pesticide use as well as energy use for irrigation and soil tillage in each treatment were recorded.

**Survey of farmers.** Representative farmers were selected for a face-to-face, questionnaire-based household survey conducted between 2007 and 2009. A total of 6,592 farmers (55 counties in 21 provinces), 6,940 farmers (113 counties in 18 provinces) and 5,406 farmers (66 counties in 22 provinces) were surveyed for rice, wheat and maize in China, respectively. In each province, three to seven counties were randomly selected, and three townships were randomly selected in each county, then two to five villages were randomly selected in each township, and finally about 20 farmers from a village were randomly surveyed to collect information on fertilizer use and grain yield in each farmer's household (Extended Data Table 1). All of these in-house surveys were conducted by professional research staff. Before beginning the survey, an informed consent information sheet was given to each farmer to read (or in some cases was read to the farmer), and verbal informed consent was requested.

**Data sources for establishing models of reactive nitrogen losses.** An exhaustive literature survey of peer-reviewed publications was undertaken using the ISI Web of Science (Thomson Reuters), Google Scholar (Google) and the China Knowledge Resource Integrated (CNKI) database, to identify articles published before December 2013. The literature survey focused on field measurements of nitrogen losses in the major Chinese agricultural regions, including NH<sub>3</sub> volatilization, nitrogen leaching, N<sub>2</sub>O emissions and nitrogen runoff. Studies had to meet specific criteria to be included in the data set. First, nitrogen losses must have been measured both during field operations and throughout the entire growing season. Second, NH<sub>3</sub> volatilization must have been measured using either the micrometeorological method<sup>27</sup> or the wind tunnel method<sup>28</sup> within at least 2 weeks after nitrogen fertilization. The N<sub>2</sub>O emissions must have been measured using the static chamber technique<sup>29</sup>, daily for 7–10 days after nitrogen fertilization and for 3–10 days after other events that may have triggered N<sub>2</sub>O gas emissions such as rainfall, irrigation or tillage, as well as weekly or biweekly during the remaining periods; and nitrogen leaching must have been measured using the suction cap or lysimeter method<sup>30</sup> or the soil sample method<sup>31</sup>. Third, only studies that reported crop yields were included. Based on the literature survey, the final data set consisted of 134 published references and 787 observations (Supplementary Information, extended reference list).

**Reactive nitrogen losses and GHG emission calculations.** Using extensive and localized databases, reactive nitrogen loss models were developed based on the relationships between N<sub>2</sub>O emission, nitrogen leaching, runoff or NH<sub>3</sub> volatilization, and nitrogen application rate or nitrogen surplus<sup>19,32</sup>. These relationships were subjected to linear or exponential regression analysis to identify the best-fit curves. Corrected R<sup>2</sup> values were used for model selection in addition to visual inspection of each response curve type. The results revealed an exponential relationship between the nitrogen surplus and direct N<sub>2</sub>O emissions, nitrogen leaching and runoff, while NH<sub>3</sub> volatilization was linearly correlated with the rate of nitrogen fertilizer application (Extended Data Figs 2–4).

Nitrogen surplus was defined as nitrogen application minus above-ground nitrogen uptake. Nitrogen uptake in field experiments was calculated by measured nitrogen concentration multiplied by measured biomass, and in the survey of farmers it was calculated by reported yield multiplied by the parameters of nitrogen required to produce a unit of grain<sup>33,34</sup>. Based on the established reactive nitrogen loss models, we calculated the amount of reactive nitrogen lost to the environment, expressed

as kilograms of nitrogen per hectare, and the reactive nitrogen loss intensity (reactive nitrogen losses per unit grain yield), expressed as kilograms of nitrogen per million grams.

The total GHG emissions, including CO<sub>2</sub>, CH<sub>4</sub> and N<sub>2</sub>O during the whole life cycle of crop production, consisted of three components: (1) those during nitrogen fertilizer application, including direct and indirect N<sub>2</sub>O emissions, which can be calculated based on the empirical reactive nitrogen losses model mentioned above; (2) those during nitrogen fertilizer production and transportation; and (3) those during the production and transportation of phosphorus and potassium fertilizer and pesticides to the farm gate, and diesel fuel use in farming operations such as sowing, tillage and harvesting.

Total N<sub>2</sub>O emissions resulting from anthropogenic nitrogen inputs to agricultural soils occur through a direct pathway (that is, directly from the soils to which the nitrogen is added), and through two indirect pathways, via the volatilization of compounds such as NH<sub>3</sub> and NO<sub>x</sub> with subsequent re-deposition downwind, and N<sub>2</sub>O emission there, and through leaching and runoff and subsequent N<sub>2</sub>O emission downstream. Indirect N<sub>2</sub>O emissions can be estimated following the IPCC methodology<sup>35</sup>, whereby 1% and 0.75% of the volatilized NH<sub>3</sub>-N and leached NO<sub>3</sub>-N are lost as N<sub>2</sub>O-N, respectively. For rice, the impact of the CH<sub>4</sub> emissions was calculated as carbon dioxide equivalents with 209 and 65 kg ha<sup>-1</sup> of emissions for single rice in south and northeast China, respectively, and 245 and 323 kg ha<sup>-1</sup> for early rice and late rice in the double rice system, respectively<sup>36</sup>. The 100-year global warming potentials of CH<sub>4</sub> and N<sub>2</sub>O are 25 and 298 times the intensity of CO<sub>2</sub> on a mass basis, respectively<sup>19</sup>. The soil CO<sub>2</sub> flux was not included as a contribution to global warming potential in our analysis: the net flux is much less than gross CO<sub>2</sub> emissions that can be measured, and net fluxes have been estimated to contribute less than 1% to the global warming potential of agriculture on a global scale<sup>37</sup>. The change of soil organic carbon content was also not included in our analysis, because it was difficult to detect small changes in the short time our experiments were in place.

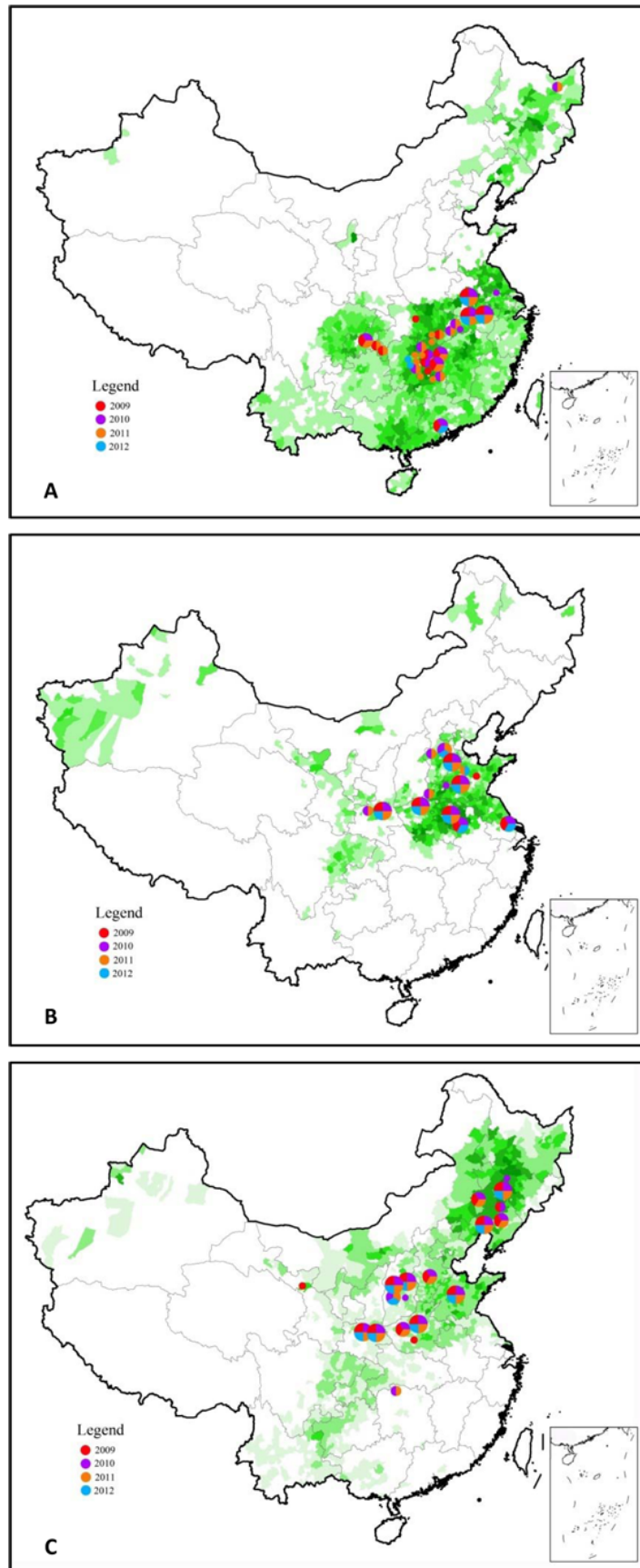
System boundaries were set as the periods of the life cycle from the production of inputs (such as fertilizers and pesticides), delivery of the inputs to the farm gates, farming operations and the crop harvesting period. Using the emission factors for all agricultural inputs given in Supplementary Table 1, we calculated total global warming potential per unit area, expressed as kilograms of carbon dioxide equivalents per hectare, and the GHG intensity, expressed as kilograms of carbon dioxide equivalents per million grams of grain.

**Projection of food demand for China in 2030.** The human population of China is projected to reach a peak of 1.47 billion around 2030, and the diet structure will change to more animal-derived protein with the development of urbanization (urbanization is projected to reach 80% in 2030)<sup>25</sup>. Using a nutrient flows in food chains, environment and resource (NUFERNUFER) model<sup>25</sup>, we project that the demand for rice, wheat and maize in 2030 for China will be 218, 125 and 315 Mt, respectively, for a total of 658 Mt for the three crops. Demand for animal feed is expected to include 308 Mt of maize and another 50 Mt of soybean.

**Data analysis.** For all field experiments, data analysis used one-way analysis of variance in SAS<sup>38</sup>. The means of management treatments were compared using least significant difference at a 0.05 level of significance for grain yield, nitrogen application, PFP<sub>N</sub>, nitrogen surplus, reactive nitrogen losses and GHG emissions.

26. Chen, X. P. *et al.* Synchronizing N supply from soil and fertilizer and N demand of winter wheat by an improved N<sub>min</sub> method. *Nutr. Cycl. Agroecosyst.* **74**, 91–98 (2006).
27. Denmead, O. T. in *Gaseous Loss of Nitrogen from Plant-Soil Systems* (eds Freney, J. R. & Simpson, J. R.) 133–157 (Martinus Nijhoff, 1983).
28. Braschkat, J., Mannheim, T., Horlacher, D. & Marschner, H. Measurement of ammonia emissions after liquid manure application: I. Construction of a wind tunnel system for measurements under field conditions. *Z. Pflanz. Bodenkunde* **156**, 393–396 (1993).
29. Holland, E. *et al.* in *Standard Soil Methods for Long-Term Ecological Research* (eds Robertson, G. P., Coleman, D. C., Bledsoe, C. S. & Sollins, P.) 185–201 (Oxford Univ. Press, 1999).
30. Lehmann, J. & Schroth, G. in *Trees, Crops and Soil Fertility: Concepts and Research Methods* (eds Schroth, G. & Sinclair, F. L.) 151–166 (CAB International, 2003).
31. Zhao, R., Chen, X. & Zhang, F. Nitrogen cycling and balance in winter wheat/summer maize rotation system in Northern China [in Chinese]. *Acta Pedol. Sin.* **46**, 684–697 (2009).
32. Grassini, P. & Cassman, K. G. High-yield maize with large net energy yield and small global warming intensity. *Proc. Natl Acad. Sci. USA* **109**, 1074–1079 (2012).
33. Hou, P. *et al.* Grain yields in relation to N requirement: optimizing nitrogen management for spring maize grown in China. *Field Crops Res.* **129**, 1–6 (2012).
34. Yue, S. C. *et al.* Change in nitrogen requirement with increasing grain yield for winter wheat. *Agron. J.* **104**, 1–7 (2012).
35. Intergovernmental Panel on Climate Change in *2006 IPCC Guidelines for National Greenhouse Gas Inventories* 11.1–11.54 (National Greenhouse Gas Inventories, 2006).
36. Feng, J. F. *et al.* Impacts of cropping practice on yield-scale greenhouse gas emission from rice fields in China: a meta-analysis. *Agric. Water Manage.* **164**, 220–228 (2013).

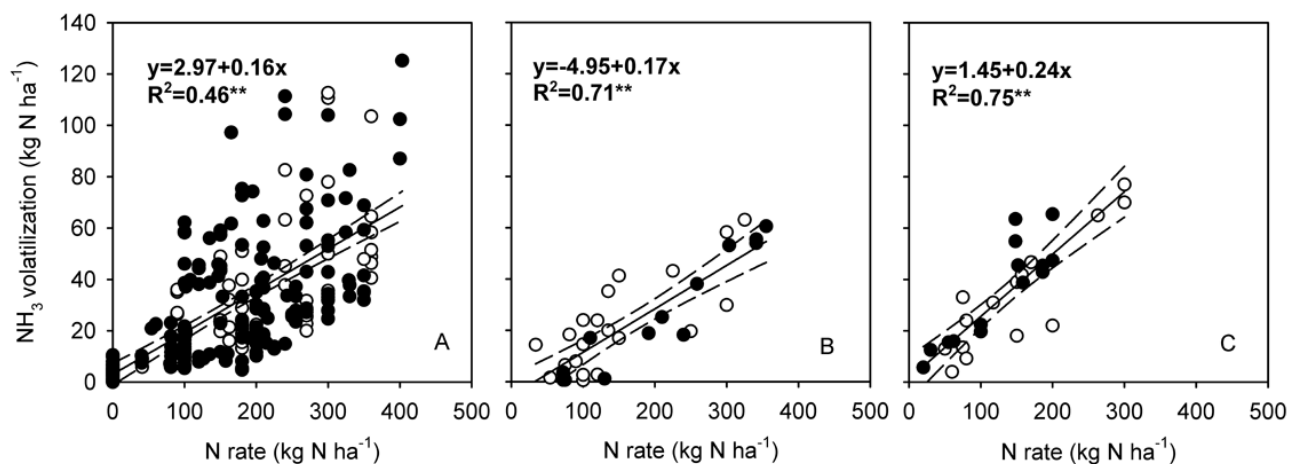
37. Smith, P. *et al.* in *Climate Change 2007: Mitigation. Contribution of Working Group III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Metz, B. *et al.*) 497–540 (Cambridge Univ. Press, 2007).
38. SAS Institute. *SAS User's Guide: Statistics* (SAS Institute, 1998).
39. National Bureau of Statistics of China. *Chinese Statistical Yearbook 2013* (China Statistics, 2013).
40. National Bureau of Statistics of China. *Nongken Agriculture Yearbook* (China Statistics, 2013).



**Extended Data Figure 1 | The distribution of experiments for grain from 2009 to 2012 in China. a, Rice ( $n = 57$ ); b, wheat ( $n = 40$ ); c, maize ( $n = 56$ ).** The background green colour represents the planting area for each crop; darker

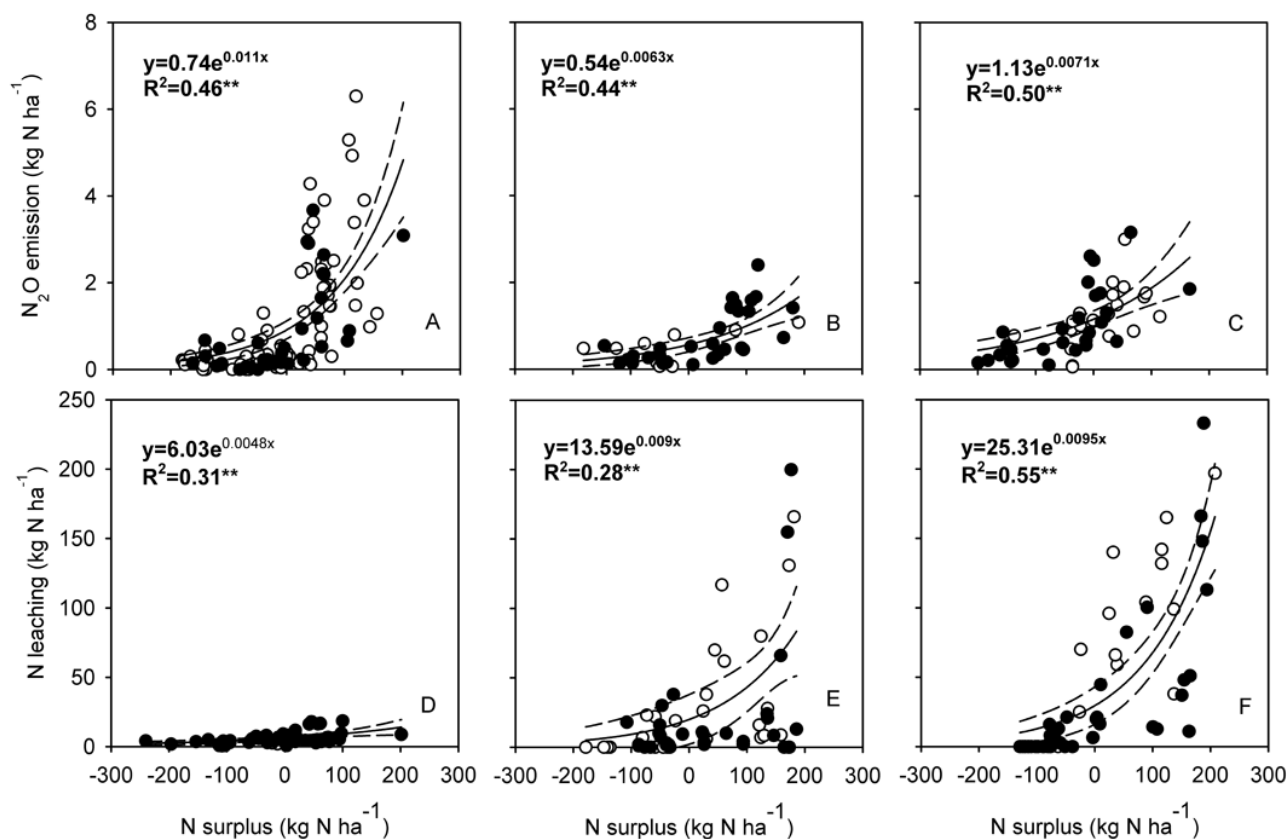
green means a larger density of planting area regionally for that crop. The dots represent sites, and each colour in a dot represents a year of measurements.





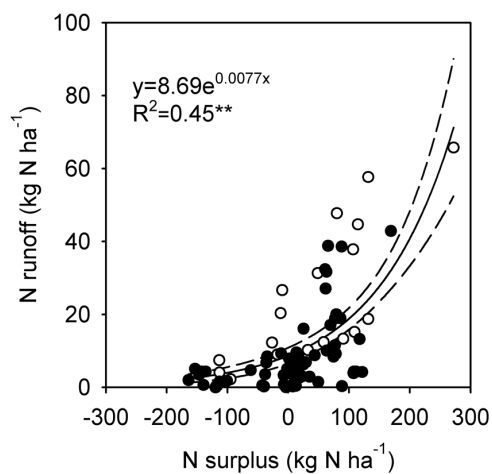
**Extended Data Figure 2 | Linear models of NH<sub>3</sub> volatilization based on nitrogen application rate.** Rate of nitrogen fertilizer application was plotted against NH<sub>3</sub>-N volatilization for (a) rice ( $n = 265$ ) (Supplementary Information, extended references 1–36 for rice), (b) wheat ( $n = 34$ ) and

(c) maize ( $n = 29$ ) (Supplementary Information, extended references 37–60 for wheat and maize) growing seasons, respectively.  $^{**}P = 0.01$ . Filled and hollow circles represent data from Chinese journals (or theses) and ISI journals, respectively.



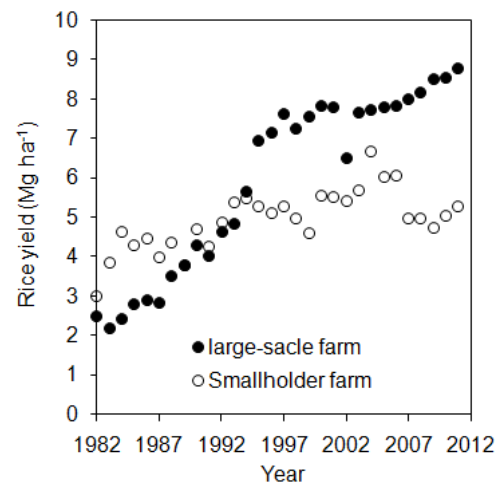
**Extended Data Figure 3 | Exponential models of N<sub>2</sub>O emissions and nitrogen leaching based on nitrogen surplus.** Nitrogen surplus was plotted against N<sub>2</sub>O-N emissions for (a) rice ( $n = 118$ ) (Supplementary information, extended references 7, 36, 61–84 for rice), (b) wheat ( $n = 40$ ) and (c) maize ( $n = 48$ ) growing seasons (Supplementary information, extended references 85–99 for wheat and maize), and against nitrogen leaching for (d) rice ( $n = 52$ )

(Supplementary information, extended references 7, 100–113 for rice), (e) wheat ( $n = 59$ ) and (f) maize ( $n = 56$ ) (Supplementary information, extended references 44, 114–121 for wheat and maize). Nitrogen surplus was defined as nitrogen application rate minus above-ground nitrogen uptake. \*\*Regression significant at  $P < 0.01$ . Solid and hollow circles represent data from Chinese journals (or theses) and ISI journals, respectively.



**Extended Data Figure 4 | Exponential model of nitrogen runoff based on nitrogen surplus for rice production.** Nitrogen surplus was defined as nitrogen application rate minus above-ground nitrogen uptake ( $n = 81$ ) (Supplementary information, extended references 8, 104, 122–134).

**\*\* $P < 0.01$ .** Solid and hollow circles represent data from Chinese journals (or theses) and ISI journals, respectively.



**Extended Data Figure 5 | Rice yields over time in smallholder and large-scale farms in Heilongjiang province from 1982 to 2011.** Large-scale farms were around 20 ha; smallholder farms were less than 2 ha. Data from refs 39, 40.



**Extended Data Table 1 | Grain yields, nitrogen application rates, calculated PFP<sub>N</sub>, nitrogen surplus, the total and the intensity of reactive nitrogen losses and GHG emissions in farmers' fields for rice (*n* = 6,592), wheat (*n* = 6,940) and maize (*n* = 5,406) in China**

Crops	Grain Yield Mg ha <sup>-1</sup>	N rate kg N ha <sup>-1</sup>	PFP <sub>N</sub> kg kg <sup>-1</sup>	N surplus kg N ha <sup>-1</sup>	Nr losses kg N Mg <sup>-1</sup>	Nr intensity kg N ha <sup>-1</sup>	GHG emission kg CO <sub>2</sub> eqha <sup>-1</sup>	GHG emission intensity kg CO <sub>2</sub> eqMg <sup>-1</sup>
Rice	7.0	209	41	82	66	9.9	10,343	1,574
	(4.5-9.8)	(86-412)	(16-83)	(-46-280)	(28-142)	(3.7-21)	(4,365-15,465)	(580-2,618)
Wheat	5.7	210	33	74	65	12	3,707	671
	(3.8-7.5)	(81-360)	(15-67)	(-49-223)	(19-147)	(2.3-27)	(2,203-5,766)	(368-1,117)
Maize	7.6	220	43	72	120	17	4,436	621
	(5.4-10.5)	(85-413)	(17-90)	(-66-256)	(36-274)	(4.6-39)	(2,273-8,269)	(287-1,179)

Values are mean and range (from the 5th to 95th percentiles).

**Extended Data Table 2 | Above-ground biomass, harvest index (HI) and crop nitrogen uptake for rice ( $n = 57$ ), wheat ( $n = 40$ ) and maize ( $n = 56$ ) in field experiments with four management treatments**

Crops	Treatment	Biomass	HI	Crop N uptake
		Mg ha <sup>-1</sup>		kg N ha <sup>-1</sup>
Rice	CP	11.7 †	0.52 *	123 ‡
	IP	12.8 *†	0.54 *	138 †
	HY	14.1 *	0.54 *	155 *
	ISSM	13.4 *†	0.54 *	147 *†
Wheat	CP	13.4 ‡	0.46 †	183 †
	IP	14.8 †‡	0.48 *†	201 †
	HY	16.7 *	0.48 *†	234 *
	ISSM	15.8 *†	0.49 *	218 *†
Maize	CP	18.4 ‡	0.49 †	194 ‡
	IP	20.8 †	0.52 *	222 †
	HY	23.7 *	0.52 *	261 *
	ISSM	23.3 *	0.52 *	249 *†

Means followed by the same letter(s) within each column for each crop are not significantly different at  $P < 0.05$ .

**Extended Data Table 3 | Yield and nitrogen rates of farmer average, top farmers and ISSM**

Crop		Yield (Mgha <sup>-1</sup> )	N rate (kg N ha <sup>-1</sup> )
Rice	Farmers average	7.0	209
	Top 20% yield of farmers	8.6	159
	ISSM	8.5	162
Wheat	Farmers average	5.7	210
	Top 5% yield of farmers	8.4	234
	ISSM	8.9	220
Maize	Farmers average	7.6	220
	Top 5% yield of farmers	11.3	229
	ISSM	14.2	256

**Extended Data Table 4 | Total production, weighted average of grain yield and nitrogen rate, total land use, nitrogen fertilizer use, reactive nitrogen losses, and GHG emissions in 2005, 2012 and projected in 2030 under three scenarios for all three crops (rice, wheat and maize) in China**

	Unit	2005	2012	2030		
				S1	S2	S3
Production	Mt	417	531	656	786	658
Grain yield	Mg ha <sup>-1</sup>	5.4	5.9	7.4	8.8	9.6
N rate	kg N ha <sup>-1</sup>	213	217	213	172	188
Land use	Million ha	78	89	89	89	69
N use	Mt	16.6	19.4	19.0	15.4	12.9
Nr losses	Mt	6.7	7.9	8.3	5.5	4.1
GHG emission	Mt CO <sub>2</sub> eq	500	558	542	498	411

Scenario 1 (S1): 'business as usual', grain yield increased by the trend with the recent 8 years (from 2005 to 2012) and nitrogen application rate the same as 2012 (there was almost no change from 2005 to 2012). Grain yields were 6.8, 5.0 and 5.9 Mg ha<sup>-1</sup>, planting areas were 30.1, 24.3 and 35.0 Mha (ref. 23), and nitrogen application rates were 209, 210, and 220 kg N ha<sup>-1</sup> in 2012 (Extended Data Table 1) for rice, wheat and maize, respectively. Scenario 2 (S2): both grain yield and nitrogen application with 80% of ISSM and land use the same as 2012. Scenario 3 (S3): both grain yield and nitrogen application with 80% of ISSM, and crop production just enough to reach projected demand for rice, wheat and maize in 2030 (requiring less cropland).



# A *Hox* regulatory network of hindbrain segmentation is conserved to the base of vertebrates

Hugo J. Parker<sup>1</sup>, Marianne E. Bronner<sup>2</sup> & Robb Krumlauf<sup>1,3</sup>

A defining feature governing head patterning of jawed vertebrates is a highly conserved gene regulatory network that integrates hindbrain segmentation with segmentally restricted domains of *Hox* gene expression. Although non-vertebrate chordates display nested domains of axial *Hox* expression, they lack hindbrain segmentation. The sea lamprey, a jawless fish, can provide unique insights into vertebrate origins owing to its phylogenetic position at the base of the vertebrate tree<sup>1–3</sup>. It has been suggested that lamprey may represent an intermediate state where nested *Hox* expression has not been coupled to the process of hindbrain segmentation<sup>4–6</sup>. However, little is known about the regulatory network underlying *Hox* expression in lamprey or its relationship to hindbrain segmentation. Here, using a novel tool that allows cross-species comparisons of regulatory elements between jawed and jawless vertebrates, we report deep conservation of both upstream regulators and segmental activity of enhancer elements across these distant species. Regulatory regions from diverse gnathostomes drive segmental reporter expression in the lamprey hindbrain and require the same transcriptional inputs (for example, *Kreisler* (also known as *Mafba*), *Krox20* (also known as *Egr2a*)) in both lamprey and zebrafish. We find that lamprey *hox* genes display dynamic segmentally restricted domains of expression; we also isolated a conserved exonic *hox2* enhancer from lamprey that drives segmental expression in rhombomeres 2 and 4. Our results show that coupling of *Hox* gene expression to segmentation of the hindbrain is an ancient trait with origin at the base of vertebrates that probably led to the formation of rhombomeric compartments with an underlying *Hox* code.

The hindbrain of jawed vertebrates is a specialized region of the nervous system characterized by its subdivision into repetitive segments called rhombomeres<sup>7</sup>. Anterior *Hox* genes are expressed in a nested pattern that is functionally coupled to this inherent segmentation program<sup>8–10</sup>. Non-vertebrate chordates possess patterned *hox* gene expression along the body axis<sup>11–14</sup>, which may be regulated by conserved patterning signals in chordate evolution<sup>15</sup>, but lack nervous system segmentation. Moreover, key segmental regulatory elements from jawed vertebrate *Hox* clusters are not conserved in amphioxus or ascidians<sup>16–18</sup>. In jawed vertebrates (gnathostomes), a well-characterized, highly conserved gene regulatory network (GRN) integrates hindbrain segmentation and *Hox* patterning<sup>8,9</sup>. The jawless (agnathan) vertebrate, lamprey, has been postulated to represent an intermediate state with rudimentary hindbrain segmentation, but lacking registration with motoneuron patterning or nested *Hox* expression<sup>4–6</sup>. However, little is known about gene regulatory events underlying *Hox* expression or coupling to hindbrain segmentation in lamprey. Here we address the nature of the agnathan hindbrain GRN and the degree to which it has been evolutionarily conserved with that of gnathostomes.

To explore upstream GRN inputs regulating *Hox* expression, we first asked whether gnathostome hindbrain regulatory elements were functional in the sea lamprey, *Petromyzon marinus*, the only agnathan for which the genome is sequenced and experimental manipulation of early embryos is feasible<sup>3,19</sup>. Using transgenic methodologies<sup>20,21</sup>, we developed a novel cross-species approach to compare activity of specific

regulatory elements between jawed and jawless vertebrates, by creating a new construct that allows efficient transgenesis in both lamprey and zebrafish embryos. We chose a series of enhancers from different jawed vertebrates that mediate segmentally restricted expression in their species of origin (Fig. 1a and Extended Data Fig. 1), focusing on elements that worked across multiple species and have well-characterized direct inputs from *Krox20*, *Kreisler*, retinoic acid and/or *Hox* auto/cross-regulation.

Analysis of F0 zebrafish embryos demonstrated that the majority of enhancers direct appropriate green fluorescent protein (GFP) reporter expression in segmental hindbrain domains (Fig. 1b, Extended Data Fig. 2 and Extended Data Table 1). The identities of segmental domains were determined by examining GFP expression in a zebrafish line expressing mCherry in rhombomere (r)3/r5 under control of the endogenous *krox20* locus. F1 lines were generated for many constructs and exhibited identical segmental expression patterns to those in F0 (Fig. 1b and Extended Data Figs 2, 3), confirming that analysis in F0 embryos accurately reflects enhancer-mediated regulatory activities.

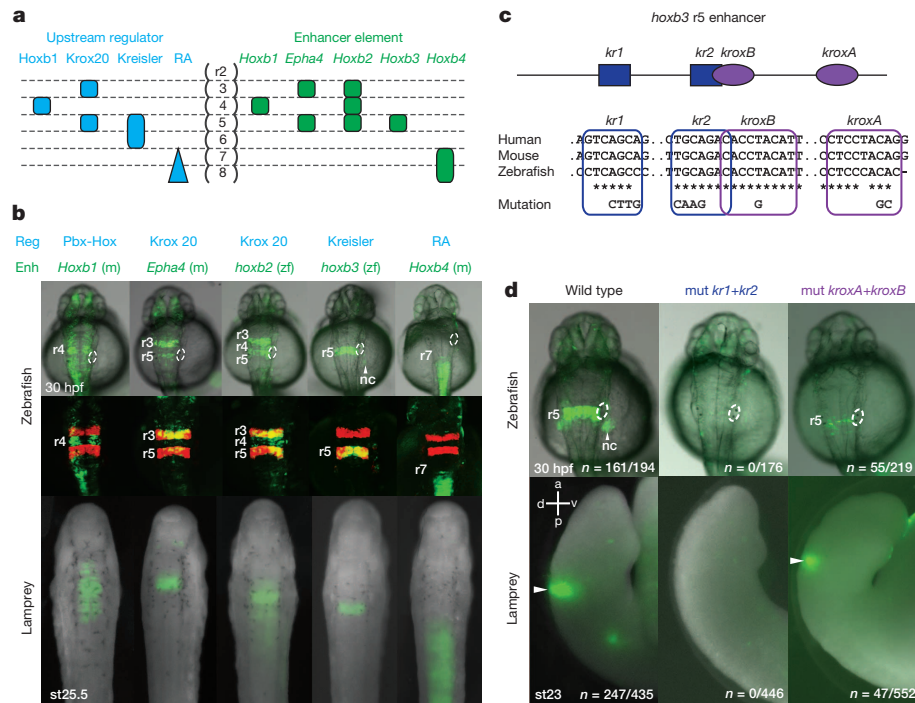
When tested for regulatory activity in lamprey, the same gnathostome constructs mediated segmental reporter expression, reminiscent of that seen in their host species and/or zebrafish. The restricted stripes of GFP expression reflect an ordered series of domains (Fig. 1b, Extended Data Fig. 2 and Extended Data Table 2), implying that these gnathostome enhancers are activated by upstream lamprey factors to mediate reporter expression in a rhombomeric fashion. Reporter expression spans multiple developmental stages with variable onset between elements (Extended Data Fig. 4). The *Hoxb1* enhancer is active first (developmental stage (st)18) in a broad domain that becomes restricted over time, followed by *hoxb2* (st21), *hoxb3* and *Hoxb4* (st22). These data suggest that a similar underlying hindbrain GRN, with temporal colinearity reminiscent of gnathostomes, may be present in lamprey.

Gnathostome rhombomeric enhancers have known *cis*-regulatory inputs: *Krox20* for *Epha4* (ref. 22) and *Hoxb2* (refs 23, 24) and *Kreisler* and *Krox20* for *Hoxb3* (refs 25, 26). We asked whether their homologues might have similar segmental regulatory functions in agnathans. To test this, we generated constructs with mutated *Kreisler* and/or *Krox20* sites within the zebrafish *hoxb3* r5 enhancer (Fig. 1c, d). Mutation of the two *Kreisler* sites (mut *kr1 + kr2*) completely eliminates reporter expression in both zebrafish and lamprey, whereas mutation of the *Krox20* sites (mut *kroxA + kroxB*) modulates levels/efficiency of expression in both species (Extended Data Tables 1 and 2). These results are consistent with roles for *Kreisler* and *Krox20* in the mouse *Hoxb3* r5 enhancer<sup>26</sup>, implying homologous roles in the lamprey hindbrain.

These data suggest that major components of the hindbrain GRN upstream of *Hox* genes are conserved in lamprey. Therefore, we characterized hindbrain expression patterns of lamprey *kreisler* and *krox20* across multiple developmental stages (st19–26) (Fig. 2). *krox20* is expressed in two stripes in a manner reminiscent of its gnathostome counterpart<sup>4</sup> (Fig. 2). We isolated a *kreisler* homologous gene that is expressed in a single stripe in the lamprey hindbrain (Fig. 2), similar to mouse *Kreisler*. The restricted expression of these key upstream regulators in lamprey supports our interpretation of their inputs to reporter activities.

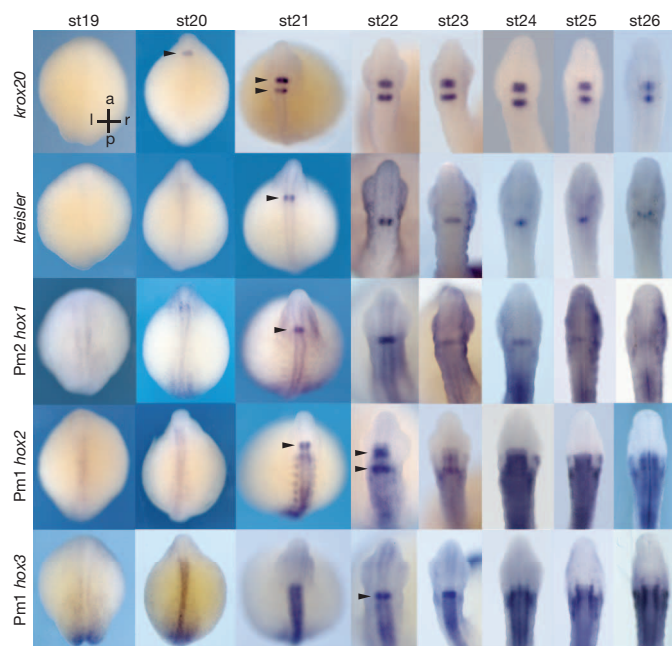
<sup>1</sup>Stowers Institute for Medical Research, Kansas City, Missouri 64110, USA. <sup>2</sup>Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, California 91125, USA.

<sup>3</sup>Department of Anatomy and Cell Biology, Kansas University Medical Center, Kansas City, Kansas 66160, USA.



**Figure 1 | Conserved segmental activity of jawed vertebrate enhancers in zebrafish and lamprey.** **a**, Schematic depicting components of the GRN for segmental *Hox* expression in the gnathostome hindbrain. The rhombomeric expression of upstream segmental regulators (blue) and the activity domains of known enhancer elements they control (green) are shown. RA, retinoic acid. **b**, GFP reporter expression in dorsal views of zebrafish and lamprey hindbrains mediated by enhancers from **a**. For zebrafish, two images of the same embryo are shown, GFP plus brightfield (top) and GFP plus endogenous r3r5-mCherry (middle) signals. The otic vesicle is circled and GFP<sup>+</sup> rhombomeres are indicated. Letters in parentheses indicate the species of origin of the element: m, mouse; zf, zebrafish. Enh, enhancer; hpf, hours post-fertilization; nc, neural

crest; Reg, regulator. **c**, The zebrafish *hoxb3* r5 enhancer contains conserved Kreisler (*kr*; blue) and Krox20 (*krox*; purple) binding sites (asterisks). Mutations known to influence activity are detailed below the aligned sites<sup>26</sup>. **d**, GFP reporter expression of wild-type and mutated (*mut*) versions of the r5 enhancer in zebrafish (dorsal views) and lamprey (lateral views) embryos. Numbers (*n*) denote the proportion of embryos exhibiting segmental reporter expression. Extended Data Tables 1 and 2 provide the number of embryos and efficiency of specific expression for all constructs. Arrowheads indicate segment-like reporter expression in the lamprey hindbrain. a, anterior; d, dorsal; p, posterior; v, ventral.

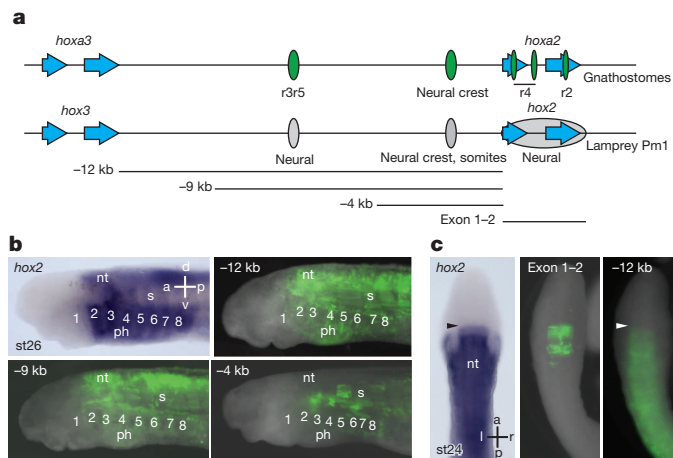


**Figure 2 | Expression of segmental regulators and *hox* genes in the lamprey hindbrain.** Gene expression visualized by *in situ* hybridization in lamprey embryos at st19–26. Dorsal views are shown, with anterior to the top. Arrowheads indicate the onset of segmental-like gene expression in the developing hindbrain. a, anterior; l, left; p, posterior; r, right.

We next examined whether lamprey *hox* genes themselves display evidence of segmental expression. We previously identified two *Hox* clusters, Pm1 and Pm2, as well as several unassigned *hox* genes in *P. marinus*<sup>3</sup>. These probably represent a subset of the total *hox* gene complement; recent evidence from *Lethenteron japonicum* suggests up to six *Hox* clusters<sup>27</sup>, two of which are homologous to Pm1 and Pm2. Lamprey *hox* genes from paralogous groups 1–3, *hox1* (Pm2), *hox2* and *hox3* (Pm1), display temporally dynamic hindbrain expression patterns. Early developmental stages (st21–23) reveal prominent stripes of restricted expression in the hindbrain for all three genes, apparently reflecting off-set segmental domains (Fig. 2) temporally correlating with robust stripes of both *krox20* and *kreisler* expression. Later (st24–26), *hox1* and *kreisler* are progressively downregulated in the hindbrain, while segmental expression for *hox2* and *hox3* become masked by their upregulation in other regions (Fig. 2). *krox20* expression initiates at st20 and remains on throughout this developmental time course. Although previous analysis of *hox* gene expression, focusing on st26 in the Japanese lamprey, found no evidence for segmental expression<sup>4,5</sup>, the potential links between *Hox* expression and hindbrain segmentation were presumably missed owing to the dynamic and early nature of segmental expression of these genes.

To identify endogenous lamprey *cis*-regulatory regions that mediate these striking segmental *hox* gene expression domains, we focused on the *hox2* paralogous group, well-characterized from a regulatory perspective in jawed vertebrates<sup>8,9</sup>. We sequenced the *hox2* locus and entire intergenic region between *hox2* and *hox3* of Pm1, as this genomic region in gnathostomes contains a series of enhancers that mediate hindbrain *Hox* expression (Fig. 3a and Extended Data Fig. 1). Because no overt sequence conservation with known jawed vertebrate enhancers was





**Figure 3 | Identification of enhancers from the lamprey *hox2* locus.** **a**, The *hoxa2-hoxa3* genomic region from gnathostomes and the equivalent region from the lamprey Pm1 *Hox* cluster. *hox* gene exons (blue arrows) and relative positions of previously characterized enhancer elements in gnathostomes (green ovals) are shown<sup>9</sup>. *hox2* enhancers identified in this study are denoted as grey ovals. Fragments of Pm1 tested in lamprey reporter assays are shown below. **b**, Lateral views of st26 lamprey embryos comparing the endogenous expression of Pm1 *hox2* with GFP reporter expression mediated by fragments of Pm1. Pharyngeal arches are numbered. nt, neural tube; s, somites; ph, pharynx. **c**, Dorsal views of st24 lamprey embryos showing endogenous expression of Pm1 *hox2* compared with GFP reporter expression. The exon 1–2 region mediates two stripes of segmental expression (Extended Data Table 2 provides information on number of embryos and efficiency of specific expression for the exon 1–2 region). Arrowheads indicate the anterior extent of expression in the neural tube.

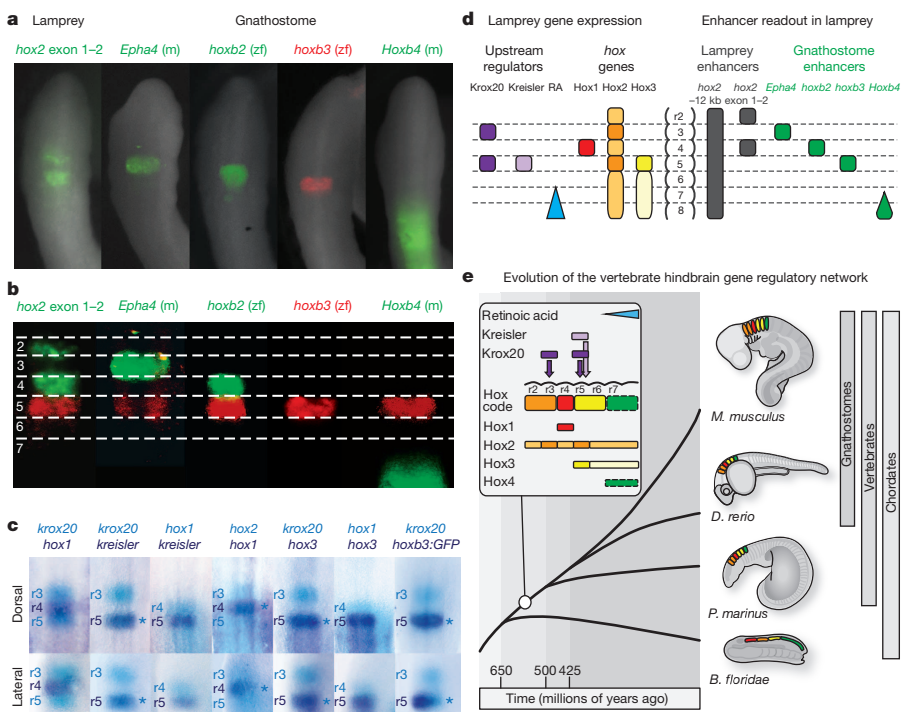
detectable, we functionally tested sequences from –12 kb upstream to +1 kb downstream of the lamprey *Hox2* coding domain in lamprey embryos (Fig. 3a–c). At st26, the –12 kb intergenic region mediates GFP expression in the neural tube, pharynx (neural crest) and somites that closely resembles that of the endogenous lamprey *hox2* gene (Fig. 3b). Deletion analyses demonstrate that *cis*-elements capable of mediating neural expression lie within the –9 kb to –4 kb intergenic region whereas

those contributing to neural crest/somite expression lie in the –4 kb fragment (Fig. 3a).

Given that gnathostome *Hoxa2* is expressed in r2 and r4 via exonic and intronic regulatory elements<sup>28–30</sup> (Fig. 3a and Extended Data Fig. 1), we tested a comparable fragment of lamprey *hox2* (exon 1–2). Intriguingly, this fragment mediated restricted expression in two alternating stripes in the hindbrain from st22 to st26 (Fig. 3c and Extended Data Fig. 4). At st24, endogenous *hox2* neural expression displays regions of varying intensity, apparently correlating with these stripes of GFP (Fig. 3c). The anterior boundary of GFP expression in the hindbrain mediated by both the –12 kb fragment and the exon 1–2 region appear to match that of the endogenous *hox2* gene (Fig. 3c). Hence, *hox2*, as in jawed vertebrates, contains multiple enhancers with partially overlapping/shadow activities. The equivalent positions of rhombomeric enhancer(s) of *hox2* and *Hoxa2* genes suggests that lamprey *hox* genes may be coupled to hindbrain segmentation in part through conserved *cis*-elements.

The lack of apparent morphological hindbrain segmentation in lamprey makes it difficult to assign these gene expression patterns to specific features. To register these expression patterns, we performed multispectral analysis using co-injection of two fluorescent reporters. The *hoxb3* enhancer was used to direct red fluorescent protein (RFP) in putative r5, allowing registration with other enhancer-mediated GFP expression (Fig. 4a, b). *hox2* exon 1–2 mediates expression in r2 and r4; *Epha4* in r3; *hoxb2* in r4; and *Hoxb4* with an anterior border of expression within r7 (Fig. 4a, b). These segmental domains generally correlate with the activity of these *cis*-elements in gnathostomes, although the *Epha4* enhancer mediates expression only in r3 in lamprey as compared with r3/r5 in zebrafish (Fig. 1b). The *hoxb2* enhancer drives robust r3/r5 expression and weaker r4 expression in zebrafish (Fig. 1b), whereas the strongest expression in lamprey is in r4 and there is weak expression in r6. Some embryos exhibit weaker r3/r5 expression, suggesting that the Krox20 sites in this enhancer are only moderately functional in lamprey. These data confirm that regulatory elements from both jawed and jawless vertebrates can mediate adjacent rhombomere-like segmental expression domains in the lamprey hindbrain.

To compare endogenous with enhancer-driven domains of expression, we performed two-colour double *in situ* hybridization (Fig. 4c). Using *krox20* as a reference for r3/r5, we mapped the site of *hox1* expression to r4. Similarly, by comparison with *krox20* and/or *hox1*, we mapped



**Figure 4 | Comparison of enhancer activity and segmental gene expression in lamprey supports an origin of the hindbrain GRN at the base of vertebrates.** **a**, **b**, The register of segmental domains of GFP expression mediated by lamprey and gnathostome enhancers in st24 lamprey embryos (**a**) are mapped to putative rhombomeres (2–7) by direct comparison with a co-injected r5 enhancer from zebrafish *hoxb3* linked to RFP (**b**).

For *hoxb2* a weaker r6 stripe begins to appear at st23. **c**, Double *in situ* hybridization reveals that endogenous *hox* gene expression and GFP reporter expression align with segmental regulators in the lamprey hindbrain. Dorsal (top) and lateral (bottom) views of st23–24 embryos are shown with anterior to the top and the inferred rhombomeric expression domains annotated. Asterisks indicate overlapping domains of *in situ* signal. **d**, Schematic summary of segmental gene expression and enhancer activity in the lamprey hindbrain at st23–24. For *hox2* and *hox3*, darker colour shades indicate stronger levels of gene expression. **e**, An evolutionary model based on our data, indicating that the GRN coupling the Hox code in the neural tube to hindbrain segmentation (rhombomeres) via Krox20 and Kreisler evolved before the split between jawed and jawless vertebrates.

*kreisler* expression domains to r5, *hox2* to r2–5 with elevated stripes in r3/r5, and the anterior stripe of *hox3* expression to r5. An antisense GFP probe positions expression directed by the *hoxb3* enhancer to r5. This analysis demonstrates that lamprey *hox* genes are expressed in a nested pattern that corresponds to the same segmental territories as their gnathostome counterparts.

By taking advantage of the unique evolutionary position of lamprey at the base of vertebrates, we have resolved a fundamental question in vertebrate evolution concerning the origin of segmental Hox patterning in the hindbrain. Our results reveal an amazing degree of conservation in both transcriptional inputs (Krox20, *Kreisler*) and regulatory element activity between jawed and jawless vertebrates (Fig. 4d). Lamprey *hox* genes display transient offset segmental expression domains, implying that the lamprey hindbrain, as in gnathostomes, is composed of identifiable rhombomeric segments with an underlying Hox code. Thus, we conclude that the coupling of *Hox* gene expression to segmentation of the hindbrain via Krox20 and *Kreisler* is an ancient vertebrate trait that evolved before the agnathan/gnathostome split (Fig. 4e).

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 15 May; accepted 31 July 2014.**

**Published online 14 September 2014.**

- Putnam, N. H. *et al.* The amphioxus genome and the evolution of the chordate karyotype. *Nature* **453**, 1064–1071 (2008).
- Shimeld, S. M. & Donoghue, P. C. Evolutionary crossroads in developmental biology: cyclostomes (lamprey and hagfish). *Development* **139**, 2091–2099 (2012).
- Smith, J. J. *et al.* Sequencing of the sea lamprey (*Petromyzon marinus*) genome provides insights into vertebrate evolution. *Nature Genet.* **45**, 415–421 (2013).
- Murakami, Y. *et al.* Segmental development of reticulospinal and branchiomotor neurons in lamprey: insights into the evolution of the vertebrate hindbrain. *Development* **131**, 983–995 (2004).
- Takio, Y. *et al.* Evolutionary biology: lamprey *Hox* genes and the evolution of jaws. *Nature* <http://dx.doi.org/10.1038/nature02616> (2004).
- Takio, Y. *et al.* Hox gene expression patterns in *Lethenteron japonicum* embryos—insights into the evolution of the vertebrate Hox code. *Dev. Biol.* **308**, 606–620 (2007).
- Lumsden, A. Segmentation and compartment in the early avian hindbrain. *Mech. Dev.* **121**, 1081–1088 (2004).
- Alexander, T., Nolte, C. & Krumlauf, R. *Hox* genes and segmentation of the hindbrain and axial skeleton. *Annu. Rev. Cell Dev. Biol.* **25**, 431–456 (2009).
- Tümpel, S., Wiedemann, L. M. & Krumlauf, R. in *HOX Genes* Vol. 88 (ed. Pourquié, O.) Ch. 4, 103–137 (Academic, 2009).
- Wilkinson, D. G., Bhatt, S., Cook, M., Boncinelli, E. & Krumlauf, R. Segmental expression of Hox-2 homeobox-containing genes in the developing mouse hindbrain. *Nature* **341**, 405–409 (1989).
- Wada, H., Garcia-Fernandez, J. & Holland, P. W. Colinear and segmental expression of amphioxus *Hox* genes. *Dev. Biol.* **213**, 131–141 (1999).
- Seo, H. C. *et al.* Hox cluster disintegration with persistent anteroposterior order of expression in *Oikopleura dioica*. *Nature* **431**, 67–71 (2004).
- Ikuta, T., Yoshida, N., Satoh, N. & Saiga, H. *Ciona intestinalis* *Hox* gene cluster: its dispersed structure and residual colinear expression in development. *Proc. Natl Acad. Sci. USA* **101**, 15118–15123 (2004).
- Lowe, C. J. *et al.* Anteroposterior patterning in hemichordates and the origins of the chordate nervous system. *Cell* **113**, 853–865 (2003).
- Pani, A. M. *et al.* Ancient deuterostome origins of vertebrate brain signalling centres. *Nature* **483**, 289–294 (2012).
- Schubert, M., Holland, N. D., Laudet, V. & Holland, L. Z. A retinoic acid-*Hox* hierarchy controls both anterior/posterior patterning and neuronal specification in the developing central nervous system of the cephalochordate amphioxus. *Dev. Biol.* **296**, 190–202 (2006).
- Natale, A. *et al.* Evolution of anterior *Hox* regulatory elements among chordates. *BMC Evol. Biol.* **11**, 330 (2011).
- Manzanas, M. *et al.* Conservation and elaboration of *Hox* gene regulation during evolution of the vertebrate head. *Nature* **408**, 854–857 (2000).
- Sauka-Spengler, T., Meulemans, D., Jones, M. & Bronner-Fraser, M. Ancient evolutionary origin of the neural crest gene regulatory network. *Dev. Cell* **13**, 405–420 (2007).
- Parker, H. J., Sauka-Spengler, T., Bronner, M. & Elgar, G. A reporter assay in lamprey embryos reveals both functional conservation and elaboration of vertebrate enhancers. *PLoS ONE* **9**, e85492 (2014).
- Parker, H. J., Piccinelli, P., Sauka-Spengler, T., Bronner, M. & Elgar, G. Ancient Pbx-*Hox* signatures define hundreds of vertebrate developmental enhancers. *BMC Genomics* **12**, 637 (2011).
- Theil, T. *et al.* Segmental expression of the *EphA4* (*Sek-1*) receptor tyrosine kinase in the hindbrain is under direct transcriptional control of Krox-20. *Development* **125**, 443–452 (1998).
- Sham, M. H. *et al.* The zinc finger gene *Krox-20* regulates *Hoxb-2* (*Hox2.8*) during hindbrain segmentation. *Cell* **72**, 183–196 (1993).
- Nonchev, S. *et al.* The conserved role of *Krox-20* in directing *Hox* gene expression during vertebrate hindbrain segmentation. *Proc. Natl Acad. Sci. USA* **93**, 9339–9345 (1996).
- Manzanas, M. *et al.* Segmental regulation of *Hoxb3* by *kreisler*. *Nature* **387**, 191–195 (1997).
- Manzanas, M. *et al.* Krox20 and *kreisler* co-operate in the transcriptional control of segmental expression of *Hoxb3* in the developing hindbrain. *EMBO J.* **21**, 365–376 (2002).
- Mehta, T. K. *et al.* Evidence for at least six *Hox* clusters in the Japanese lamprey (*Lethenteron japonicum*). *Proc. Natl Acad. Sci. USA* **110**, 16044–16049 (2013).
- Tümpel, S., Cambrero, F., Sims, C., Krumlauf, R. & Wiedemann, L. M. A regulatory module embedded in the coding region of *Hoxa2* controls expression in rhombomere 2. *Proc. Natl Acad. Sci. USA* **105**, 20077–20082 (2008).
- Tümpel, S., Cambrero, F., Wiedemann, L. M. & Krumlauf, R. Evolution of *cis* elements in the differential expression of two *Hoxa2* coparalogous genes in pufferfish (*Takifugu rubripes*). *Proc. Natl Acad. Sci. USA* **103**, 5419–5424 (2006).
- Tümpel, S. *et al.* Expression of *Hoxa2* in rhombomere 4 is regulated by a conserved cross-regulatory mechanism dependent upon *Hoxb1*. *Dev. Biol.* **302**, 646–660 (2007).

**Acknowledgements** We thank S. Green for sharing expertise, methods and managing the lamprey facility; T. Sauka-Spengler for *in situ* hybridization advice; M. Simoes-Costa and B. Uy for help with reporter constructs; J. McEllin, C. Nolte, C. Scott and L. Wiedemann for discussions and assistance with lamprey *hox* genes; M. Distel and R. Koster for the r3/r5-mCherry zebrafish line; M. Miller for graphic design; A. Ikmi for manuscript comments and the Stowers Institute aquatics facility for zebrafish care. H.J.P. and R.K. were supported by the Stowers Institute (RK grant #2013-1001). M.E.B. was supported by grants R01NS086907 and R01DE017911.

**Author Contributions** H.J.P., R.K. and M.E.B. conceived this research program. H.J.P. conducted the experiments. H.J.P., R.K. and M.E.B. jointly analysed the data, discussed the ideas and interpretations and wrote the manuscript.

**Author Information** The sequences for the lamprey *hox1w* and *kreisler* transcripts have been deposited in GenBank under accession numbers KM087087 (*hox1w*) and KM087088 (*kreisler*). All original source data have been deposited in the Stowers Institute Original Data Repository and are available online at <http://odr.stowers.org/websimr/>. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.K. ([rek@stowers.org](mailto:rek@stowers.org)).



## METHODS

**Enhancer elements.** Enhancer elements were selected from the published data or identified based upon cross-species sequence alignments. DNA containing each element was amplified by PCR from genomic DNA templates using Phusion High-Fidelity DNA Polymerase (NEB). The primers listed below were used for amplification and the size of each amplified fragment is indicated (in bp). The sequences in bold represent homology to genomic DNA and adaptor sequences are in non-bolded text. Mouse *Hoxb1* (ref. 31) (378 bp): F: 5'-AATTTGGGGCCCTCTAATAATCC AAGAACCCTATTGAAGG-3'; R: 5'-TACAACCTCGAGCAGTATGTCACAG AGCTGAAG-3'. Mouse *Hoxa2* (ref. 32) (808 bp) F: 5'-GATGCTGGGCCAGAG TCTGAATGCTGGAGCAGTCTCAG-3'; R: 5'-CATAGCCTCGAGGTACCT TCTCTCCCTCAAACC-3'. Zebrafish *hoxa2b* (2,960 bp) F: 5'-GGGTATTAAC AGGTATCTGAATGC-3'; R: 5'-AAATTCGCCGCTCTCAAAT-3'. *Fugu rubripes Hoxa2a*<sup>29</sup> (1,404 bp) F: 5'-ATCTGAGGGCCCTGGCTTAATGCAACGCTA TATTT-3'; R: 5'-GTACATCTCGAGCCCTATTTGCAATACGACTCTG-3'. *Fugu rubripes Hoxa2b*<sup>29</sup> (1,263 bp) F: 5'-TGCTGTAATGCCAAACCTC-3'; R: 5'-CCTGCTCGCCTTCGTGCGC-3'. Mouse *Hoxb2* (ref. 33) (2,021 bp) F: 5'-ATGCGTGGGGCCGGATCCCCACTTTAAACCCCAAG-3'; R: 5'-GTACAG CTCGAGTCTCCGCCAATCGCTAGT-3'. Zebrafish *hoxb2a* (1,488 bp) F: 5'-T GACCCCATTCGCTAGTACC-3'; R: 5'-TATTTTGCGCTCTGCTATG-3'. Mouse *Epha4* (ref. 22) (496 bp) F: 5'-AACTGAGGCGCCAGCATGGAGCTCTC TTAGCGTA-3'; R: 5'-TCATTAATCTCGAGTTTCGGGCTCTAGATCTGC-3'. Mouse *Hoxb3* (ref. 25) (649 bp) F: 5'-AGCTCTCTCGAGCAGTAGGATCCCAG GT-3'; R: 5'-GCTAATCTCGAGGAGGCGCTGTAGGAGAAG-3'. Zebrafish *hoxb3a* (928 bp) F: 5'-AATGGAGGGCCGTTGCCGAAGTGTCTGTTTC-3'; R: 5'-AGGGAACTCGAGCTCCAGTGTCTGCTGTC-3'. Mouse *Hoxb4* (ref. 34) (951 bp) F: 5'-AACTGAGGGCCCTGGAATTTGGTTGGGTTTCT-3'; R: 5'-TA TCTCTCGAGTGTCCATGGTGGAAAGC-3'. Mouse *Hoxd4* (ref. 35) (582 bp) F: 5'-ACAAGTGGGCCCTGGAGGAAGGCTAGCTTAA-3'; R: 5'-AAAAA GCTCGAGAAGGGTAGTTAAAGTCCAAAGG-3'. Lamprey *hox2* exon 1–2 (2,625 bp) F: 5'-CGATGAGTCGACAGTTTGAGCGGGAACTGG-3'; R: 5'-CTAATCGTCGACCGAAATCTATTGCGCCTACA-3'.

**Generation of reporter constructs.** The Hugo's lamprey construct (HLC) vector and its variants (*HLC-GW*, *HLC-RFP*) were created for this study (reagents and sequences are available on request). PCR-purified enhancer elements were cloned into HLC using either standard restriction-enzyme-mediated methods or by first cloning PCR products into the *pCR8/GW/TOPO TA* vector (Invitrogen) followed by transfer into a Gateway-compatible variant of HLC (*HLC-GW*) via *in vitro* recombination using the Gateway LR-Clonase II enzyme (Invitrogen).

The 12 kb intergenic region between lamprey *hox2* and *hox3* of the Pm1 cluster was cloned into HLC by homologous capture from lamprey bacterial artificial chromosome (BAC) 218A09 (L6)<sup>3</sup> following previously described recombineering methods<sup>36</sup> and using the following homology arm sequences (homology arms indicated in bold). Arm 1, 5'-GGGCCGTACACGGACCTGTCGTCTCATCACC ACCGACTCAGGAAGTACTAGT-3'; arm 2, 5'-ACACCCCCCCCCCTCTCGCTCAGTGCTCCGTCAGGCAGCCATGG-3'.

Shorter fragments of this intergenic region were subsequently generated from the captured 12 kb sequence by standard restriction-enzyme-mediated cloning approaches.

Site-directed mutagenesis was performed on the zebrafish *hoxb3* HLC construct using the QuikChange II XL Site-Directed Mutagenesis Kit (Stratagene) and the primers listed below. The bold text indicates mutated sequences that differ from wild type. *kr1mutF*: 5'-GTGTTTTCTGCATTTCTGTCCTCTTGACGCTG TTAGTTAATTAGTG-3'; *kr1mutR*: 5'-CACTAATTAACAAACAGCTGCAAG GAGGCAACGAAATGCAGAAACAC-3'; *kr2mutF*: 5'-CAATGCCGTTTATG TAAAAAGTCAAGGACACCTACATTTTGCCTTG-3'; *kr2mutR*: 5'-CAAGGC AAAAATGTAGGTGCTTGACTTTTACTAAACGGCATTG-3'; *kroxAmutF*: 5'-GCCTTCTCTCCAGCCCCGTTGGTGATGC-3'; *kroxAmutR*: 5'-GCATCAC CAACGGGCTGGGAGGAAGGC-3'; *kroxBmutF*: 5'-GTTGCAGACCCGAC ATTTTTCCTTGTGC-3'; *kroxBmutR*: 5'-GCACAAGGCAAAATGTCTCGGT GTCTGCAAC-3'.

**Zebrafish and lamprey experiments.** This study was conducted in accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health and protocols were approved by the Institutional Animal Care and Use Committees of the Stowers Institute (zebrafish, RK Protocol #2013-0110) and California Institute of Technology (lamprey, MEB Protocol #1436-11).

**Zebrafish reporter assay.** The following zebrafish lines were used for embryo micro-injection experiments: Slusarski AB (wild type); *egr2b:KalTA4B1-1xUASCherry* (r3r5-mCherry)<sup>37</sup>. Transient transgenic zebrafish embryos were generated for each reporter construct by Tol2-mediated transgenesis in fertilized eggs as described

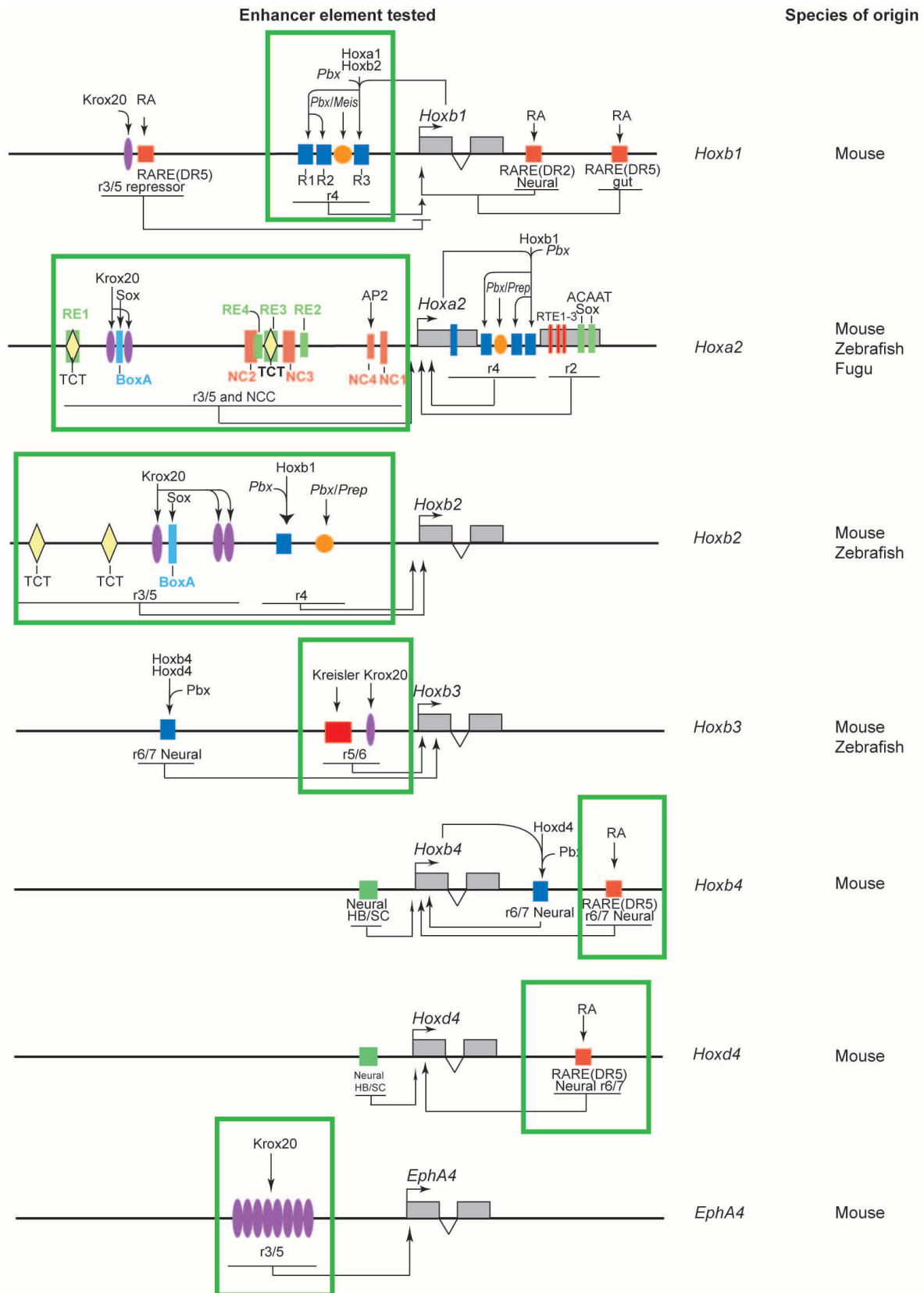
previously<sup>38</sup>. In general a minimum of 100 embryos were injected to monitor efficiency for each construct due to mosaicism and position effects of integration. GFP-expressing transient transgenic embryos were raised to adulthood and crossed with either wild-type or r3r5-mCherry fish to screen for germline transgene integration<sup>38</sup>. Embryos were screened for fluorescent reporter expression using a Leica M205FA microscope. Fluorescence and bright-field signals were captured with a Leica DFC360FX camera using LAS AF imaging software. Images were cropped and alterations to brightness and contrast were made using Adobe Photoshop CS5.1.

**Lamprey reporter assay.** Embryos were harvested from gravid lamprey (*P. marinus*) caught in the wild and provided by Hammond Bay Biological Station. Transient transgenic *P. marinus* embryos were generated by I-SceI meganuclease-mediated transgenesis as described previously<sup>20</sup>. Single-celled embryos at 4–6 h post-fertilization were injected with the digested construct at a concentration of 20 ng  $\mu\text{L}^{-1}$ , maintained as described previously<sup>19</sup> and screened for reporter expression daily from ref. 39 st17 onwards. In general a minimum of 100 embryos were injected to monitor efficiency for each construct due to mosaicism and position effects of integration. The zebrafish *hoxb3-HLC-RFP* construct, containing RFP rather than GFP, was created for the co-injection experiments. Co-injected constructs were mixed at a concentration of 15 ng  $\mu\text{L}^{-1}$  each (resulting in a total DNA concentration of 30 ng  $\mu\text{L}^{-1}$ ) and digested for injection. Embryos were screened for fluorescence using a Zeiss SteREO Discovery V12 microscope and imaged with a Zeiss Axiocam MRm camera and AxioVision Rel 4.6 software. Images were cropped and altered for brightness and contrast using Adobe Photoshop CS5.1.

**Cloning lamprey *in situ* hybridization probes.** Exonic probes were designed based on previously characterized/predicted gene sequences<sup>3</sup> and were amplified from *P. marinus* genomic DNA by PCR using Phusion High-Fidelity DNA Polymerase and cloned into the *pCR4-TOPO* vector. The size of each amplified fragment is indicated (in bp). For generating 5' and 3' untranslated region (UTR) probes, RNA from st18–26 *P. marinus* embryos was extracted using the RNeasy Total RNA Isolation Kit (Ambion) and used as a template for 5' or 3' rapid amplification of cDNA ends (RACE) with the GeneRacer Kit and SuperScript III RT (Invitrogen). cDNA fragments were amplified by PCR using Phusion High-Fidelity DNA Polymerase and cloned into the *pCR4-TOPO* vector. The following primers were used for PCR. *krox20* (ref. 4) (468 bp, predicted exonic fragment) F: 5'-CCACAAG CCCTTCAGTG-3'; R: 5'-GGTGAGGACATACGCGAGAG-3'. *kreisler* (529 bp, 5' UTR and partial exon) F: Generacer 5' Nested Primer; R: 5'-GAGAGGGCCG CTCGGAGAAGCTTGA-3'. Pm2 *hox1* (949 bp, partial exon and 3' UTR) F: 5'-CAGAACCAGGCGCATGAAGCAGAAGA-3'; R: Generacer 3' Nested Primer. Pm1 *hox2* (471 bp, partial exon 2) F: 5'-CAAGCGGCAGACTCAGTACA-3'; R: 5'-AGGTCAGCGTGCTCTCTAA-3'. Pm1 *hox3* (661 bp, partial exon 2) F: 5'-G ACGAGTTGAAATGGCCAAC-3'; R: 5'-TGAGACGACAGGTCCGTGTA-3'. *eGFP* (709 bp) F: 5'-CAAGGGCGAGGAGCTGTT-3'; R: 5'-CTTGTACAGCTC GTCCATGC-3'.

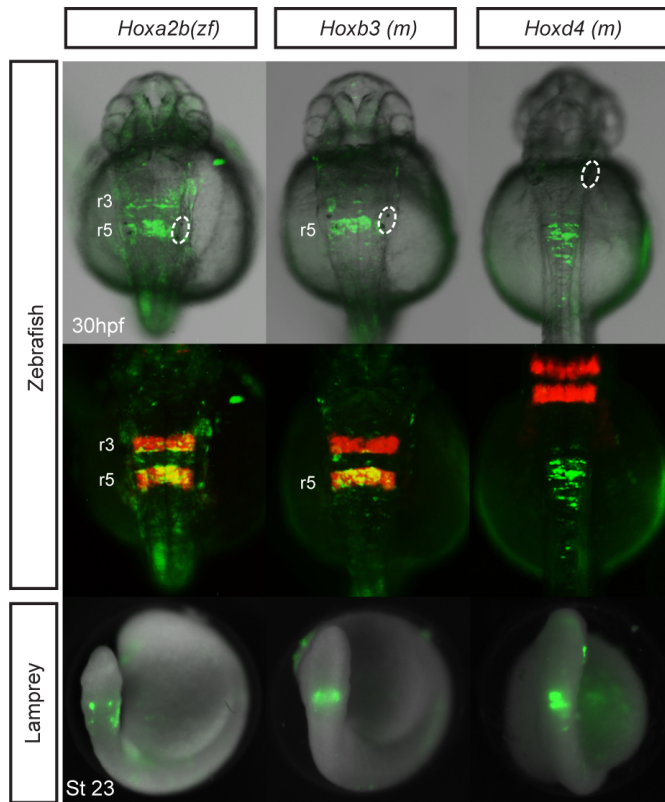
**Lamprey *in situ* hybridization.** Digoxigenin- and fluorescein-labelled probes were generated by standard methods and used in single and double lamprey whole-mount *in situ* hybridization as described previously<sup>19</sup>. Embryos were cleared in a solution of 75% glycerol before being imaged using a Leica MZ APO microscope with a Zeiss Axiocam HRC camera and Axiovision Rel 4.8 software. Images were cropped and altered for brightness and contrast using Adobe Photoshop CS5.1.

- Pöpperl, H. *et al.* Segmental expression of *Hoxb1* is controlled by a highly conserved autoregulatory loop dependent upon *exd/Pbx*. *Cell* **81**, 1031–1042 (1995).
- Nonchev, S. *et al.* Segmental expression of *Hoxa-2* in the hindbrain is directly regulated by *Krox-20*. *Development* **122**, 543–554 (1996).
- Maconochie, M. K. *et al.* Cross-regulation in the mouse *HoxB* complex: the expression of *Hoxb2* in rhombomere 4 is regulated by *Hoxb1*. *Genes Dev.* **11**, 1885–1895 (1997).
- Gould, A., Itasaki, N. & Krumlauf, R. Initiation of rhombomeric *Hoxb4* expression requires induction by somites and a retinoid pathway. *Neuron* **21**, 39–51 (1998).
- Zhang, F., Nagy Kovacs, E. & Featherstone, M. S. Murine *Hoxd4* expression in the CNS requires multiple elements including a retinoic acid response element. *Mech. Dev.* **96**, 79–89 (2000).
- Nolte, C., Jinks, T., Wang, X., Martinez Pastor, M. T. & Krumlauf, R. Shadow enhancers flanking the *HoxB* cluster direct dynamic *Hox* expression in early heart and endoderm development. *Dev. Biol.* **383**, 158–173 (2013).
- Distel, M., Wullimann, M. F. & Koster, R. W. Optimized Gal4 genetics for permanent gene expression mapping in zebrafish. *Proc. Natl Acad. Sci. USA* **106**, 13365–13370 (2009).
- Fisher, S. *et al.* Evaluating the biological relevance of putative enhancers using Tol2 transposon-mediated transgenesis in zebrafish. *Nature Protocols* **1**, 1297–1305 (2006).
- Tahara, Y. Normal stages of development in the lamprey, *Lampetra reissneri* Dybowski. *Zool. Sci.* **5.1**, 109–118 (1988).

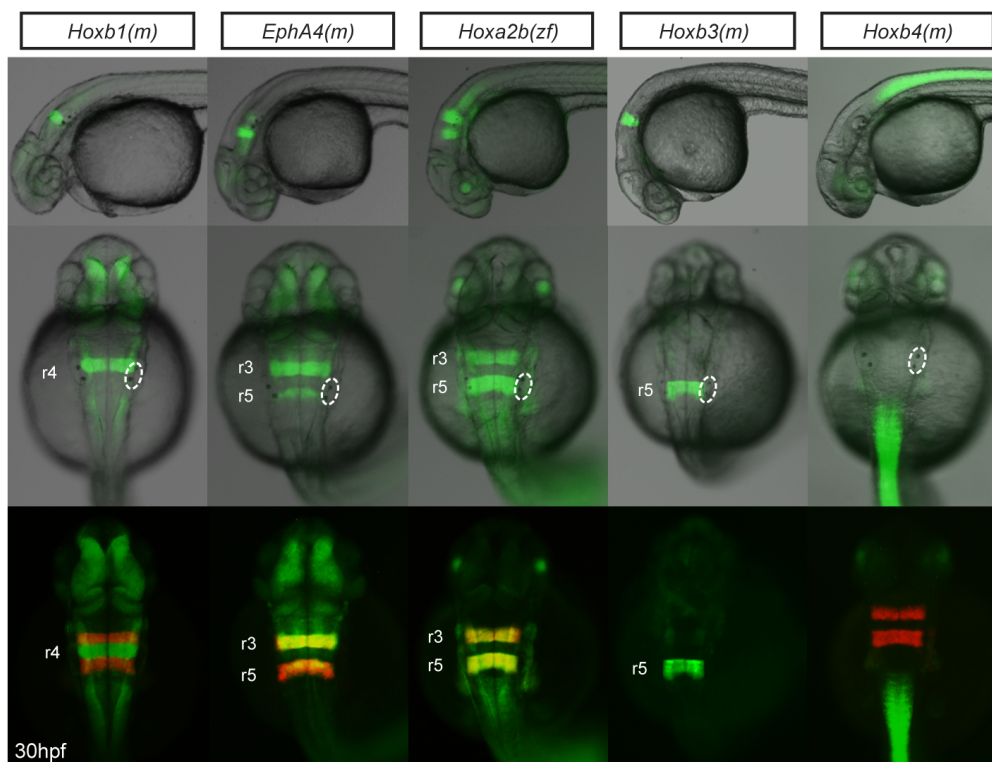


**Extended Data Figure 1 | Gnathostome enhancer elements selected for reporter analysis.** Schematic diagrams depicting the gnathostome enhancer elements assayed for activity in zebrafish and lamprey embryos in this study. The endogenous genomic positions of the enhancer elements (green boxes) are shown relative to the genes that they regulate. Known *trans*-acting factors are

listed above the elements, while the corresponding regulatory modules and their combined activity domains are detailed below the elements. For each element, the species from which it was cloned are listed on the right. Figure adapted with permission from figure 4.2 in ref. 9.



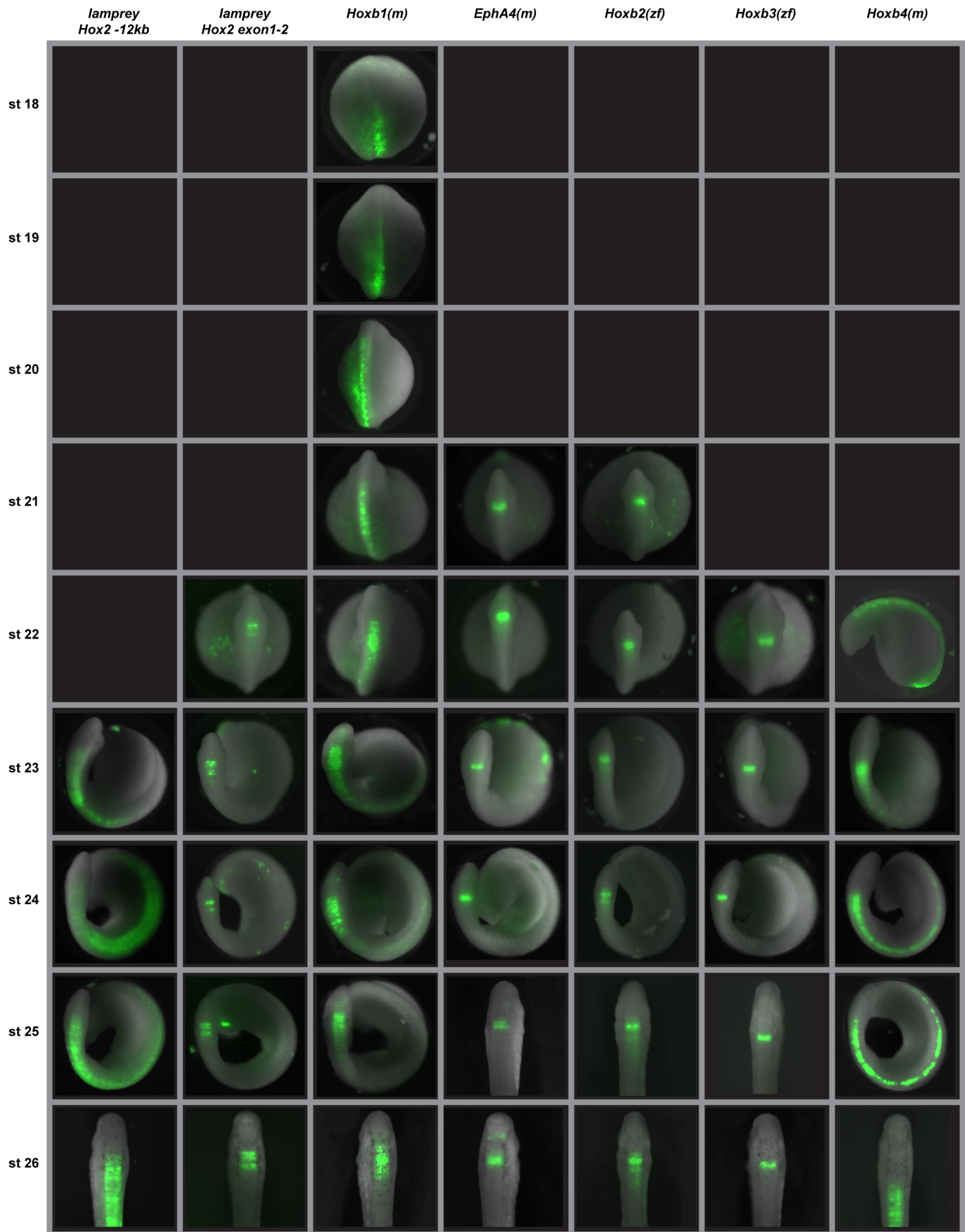
**Extended Data Figure 2 | Segmental activity of additional jawed vertebrate enhancers in zebrafish and lamprey.** GFP reporter expression mediated by gnathostome enhancer elements in zebrafish and lamprey embryos. Dorsal views are shown, with anterior to the top. For zebrafish, two images of the same embryo are shown, presenting GFP plus brightfield (top) and GFP plus endogenous r3r5-mCherry (middle) signals. The zebrafish otic vesicle is circled. m, mouse; zf, zebrafish.



**Extended Data Figure 3 | Segmental patterns of GFP reporter expression in transgenic zebrafish lines.** Lateral (top) and dorsal (middle) views of 30 hpf transgenic (F1) zebrafish embryos show combined brightfield illumination and segmental GFP reporter expression in the hindbrain mediated by five different gnathostome enhancer elements. The corresponding transient

transgenic GFP expression patterns mediated by these elements are shown in Fig. 1b and Extended Data Fig. 2. When available, GFP lines were crossed with the endogenous r3r5-mCherry reporter line as a reference (bottom). The otic vesicle is circled. m, mouse; zf, zebrafish.





**Extended Data Figure 4 | Developmental time course of GFP reporter expression mediated by lamprey and gnathostome regulatory elements in lamprey embryos.** Developmental stages st18–26 are shown. All embryos are positioned such that the hindbrain is viewed dorsally, with anterior to the top, except for mouse *Hoxb4* at st22, which is viewed laterally with anterior to the left. For *hoxb2* a weaker r6 stripe begins to appear at st23. Black boxes indicate no GFP expression mediated by that element at that developmental

stage. In both fish and lamprey, expression driven by the gnathostome *Hoxb1* enhancers appears to be temporally dynamic, starting broad and refining with time, which is probably caused by autoregulation within this element. However, we cannot rule out the possibility that the enhancers used may be missing some repressor elements that are required for fine-tuning. m, mouse; zf, zebrafish.

Extended Data Table 1 | Zebrafish reporter assay statistics

	Element	Expression domain	# embryos	# specific expression	% specific expression
Gnathostome elements	<i>Hoxb1(m)</i>	hindbrain	230	218	94.8
	<i>Hoxa2(m)</i>	hindbrain	145	64	44.1
	<i>Hoxa2b(zf)</i>	hindbrain	125	46	36.8
	<i>Hoxa2a(fr)</i>	hindbrain	123	16	13.0
	<i>Hoxa2b(fr)</i>	no specific expression	N/A	N/A	N/A
	<i>Hoxb2(m)</i>	hindbrain	199	75	37.7
	<i>Hoxb2a(zf)</i>	hindbrain	147	95	64.6
	<i>EphA4(m)</i>	hindbrain	195	172	88.2
	<i>Hoxb3(m)</i>	hindbrain	98	70	71.4
	<i>Hoxb3a(zf)</i>	hindbrain	549	503	91.6
	<i>Hoxb4(m)</i>	spinal cord	160	125	78.1
	<i>Hoxd4(m)</i>	spinal cord	324	141	43.5
<i>Hoxb3a(zf)</i> dissection	<i>Hoxb3a(zf) exp 1</i>	hindbrain	194	161	83.0
	<i>Hoxb3a(zf) exp 2</i>	hindbrain	142	137	96.5
	<i>Hoxb3a(zf) exp 3</i>	hindbrain	213	205	96.2
	<i>Hoxb3a(zf) kr12 mut exp 1</i>	hindbrain	176	0	0.0
	<i>Hoxb3a(zf) kr12 mut exp 2</i>	hindbrain	220	0	0.0
	<i>Hoxb3a(zf) kroxAB mut exp 1</i>	hindbrain	162	14	8.6
	<i>Hoxb3a(zf) kroxAB mut exp 2</i>	hindbrain	219	55	25.1

For each injected construct, the tissue-specific GFP expression domains are noted, along with the number and proportion of screened embryos exhibiting GFP expression in those domains. In each case, the numbers derive from individual rounds of injection, except for zebrafish *hoxb3a*, for which the data from three separate experiments (exp 1–3), which were performed to ensure reproducibility, were combined. Letters in parentheses after the element names indicate the species of origin of the element: fr, *Fugu rubripes*; m, mouse; zf, zebrafish. N/A, numbers on efficiency not available.

Extended Data Table 2 | Lamprey reporter assay statistics

	Element	Stage	Expression domain	# embryos	# specific expression	% specific expression
Gnathostome elements	<i>Hoxb1(m)</i>	22	neural tube	231	137	59.3
	<i>Hoxa2(m)</i>	24	neural crest	264	138	52.3
	<i>Hoxa2b(zf)</i>	23	neural crest	261	120	46.0
	<i>Hoxa2a(fr)</i>	23	hindbrain and neural crest	246	57	23.2
	<i>Hoxa2b(fr)</i>	24	pharynx	218	70	32.1
	<i>Hoxb2(m)</i>	N/A	no specific expression	N/A	N/A	N/A
	<i>Hoxb2a(zf)</i>	23	hindbrain	192	113	58.9
	<i>EphA4(m)</i>	23	hindbrain	695	100	14.4
	<i>Hoxb3(m)</i>	24	hindbrain	324	32	9.9
	<i>Hoxb3a(zf)</i>	23	hindbrain	1440	474	32.9
	<i>Hoxb4(m)</i>	24	hindbrain & spinal cord	590	169	28.6
	<i>Hoxd4(m)</i>	25	hindbrain & spinal cord	300	28	9.3
<i>Hoxb3a(zf)</i> dissection	<i>Hoxb3a(zf) exp 1</i>	23	hindbrain	435	247	56.8
	<i>Hoxb3a(zf) exp 2</i>	23	hindbrain	557	93	16.7
	<i>Hoxb3a(zf) exp 3</i>	23	hindbrain	448	134	29.9
	<i>Hoxb3a(zf) kr12 mut exp 1</i>	23	hindbrain	407	0	0.0
	<i>Hoxb3a(zf) kr12 mut exp 2</i>	23	hindbrain	437	2	0.5
	<i>Hoxb3a(zf) kr12 mut exp 3</i>	23	hindbrain	446	0	0.0
	<i>Hoxb3a(zf) kroxAB mut</i>	23	hindbrain	522	47	9.0
Lamprey elements	<i>Hox2 -12kb</i>	24	neural tube, pharynx, somites	N/A	N/A	N/A
	<i>Hox2 -9kb</i>	23	neural tube, pharynx, somites	N/A	N/A	N/A
	<i>Hox2 -4kb</i>	23	pharynx, somites	N/A	N/A	N/A
	<i>Hox2 exon1-2</i>	23	hindbrain	406	123	30.3

For each injected construct, the tissue-specific GFP expression domains are given, along with the number and proportion of screened embryos exhibiting GFP expression in those domains. In each case, the numbers derive from individual rounds of injection, except for zebrafish *hoxb3a*, for which the data from three separate experiments (exp 1–3), which were performed to ensure reproducibility, were combined. Letters in parentheses after the element names indicate the species of origin of the element: fr, *Fugu rubripes*; m, mouse; zf, zebrafish. N/A, numbers on efficiency not available.



# Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis

Kirsten I. Bos<sup>1\*</sup>, Kelly M. Harkins<sup>2\*</sup>, Alexander Herbig<sup>1,3\*</sup>, Mireia Coscolla<sup>4,5\*</sup>, Nico Weber<sup>3</sup>, Iñaki Comas<sup>6,7</sup>, Stephen A. Forrest<sup>1</sup>, Josephine M. Bryant<sup>8</sup>, Simon R. Harris<sup>8</sup>, Verena J. Schuenemann<sup>1</sup>, Tessa J. Campbell<sup>9</sup>, Kerttu Majander<sup>1</sup>, Alicia K. Wilbur<sup>2</sup>, Ricardo A. Guichon<sup>10</sup>, Dawnie L. Wolfe Steadman<sup>11</sup>, Della Collins Cook<sup>12</sup>, Stefan Niemann<sup>13,14</sup>, Marcel A. Behr<sup>15</sup>, Martin Zumarraga<sup>16</sup>, Ricardo Bastida<sup>17</sup>, Daniel Huson<sup>3</sup>, Kay Niselt<sup>3</sup>, Douglas Young<sup>18,19</sup>, Julian Parkhill<sup>8</sup>, Jane E. Buikstra<sup>2</sup>, Sebastien Gagneux<sup>4,5</sup>, Anne C. Stone<sup>2</sup> & Johannes Krause<sup>1,20,21</sup>

**Modern strains of *Mycobacterium tuberculosis* from the Americas are closely related to those from Europe, supporting the assumption that human tuberculosis was introduced post-contact<sup>1</sup>. This notion, however, is incompatible with archaeological evidence of pre-contact tuberculosis in the New World<sup>2</sup>. Comparative genomics of modern isolates suggests that *M. tuberculosis* attained its worldwide distribution following human dispersals out of Africa during the Pleistocene epoch<sup>3</sup>, although this has yet to be confirmed with ancient calibration points. Here we present three 1,000-year-old mycobacterial genomes from Peruvian human skeletons, revealing that a member of the *M. tuberculosis* complex caused human disease before contact. The ancient strains are distinct from known human-adapted forms and are most closely related to those adapted to seals and sea lions. Two independent dating approaches suggest a most recent common ancestor for the *M. tuberculosis* complex less than 6,000 years ago, which supports a Holocene dispersal of the disease. Our results implicate sea mammals as having played a role in transmitting the disease to humans across the ocean.**

*Mycobacterium tuberculosis* has had a long history with humans, although consensus has not been reached on when this interaction began<sup>1,3,4</sup>. Previous models held that the human-adapted pathogen evolved from a zoonotic transfer of *Mycobacterium bovis* following animal domestication during the Neolithic age<sup>5</sup>. Comparative genomic analyses, however, suggest that the bovine form and those adapted to other animal hosts are in fact derived from human strains<sup>3,6</sup>. This supports a rather different disease history where humans may have been the most susceptible host species for early progenitors of strains currently circulating. Today the majority of *M. tuberculosis* diversity exists in Africa<sup>7</sup>, implying that the pathogen probably originated from a monoclonal expansion therein and achieved its worldwide distribution via human movements<sup>1,3,4</sup>. The observations that *M. tuberculosis* strains tend to be associated with human populations<sup>8</sup> and that selection in the bacterium exists at loci associated with host immune responses<sup>9</sup> indicate that host and pathogen have had sufficient time to co-evolve. Dating approaches that use human demographic events for calibration generate substitution rates that differ by over an order of magnitude depending on the model<sup>3,10</sup>, and would thus contribute to vastly different coalescence estimates for

all *M. tuberculosis* lineages, collectively referred to as the *M. tuberculosis* complex (MTBC).

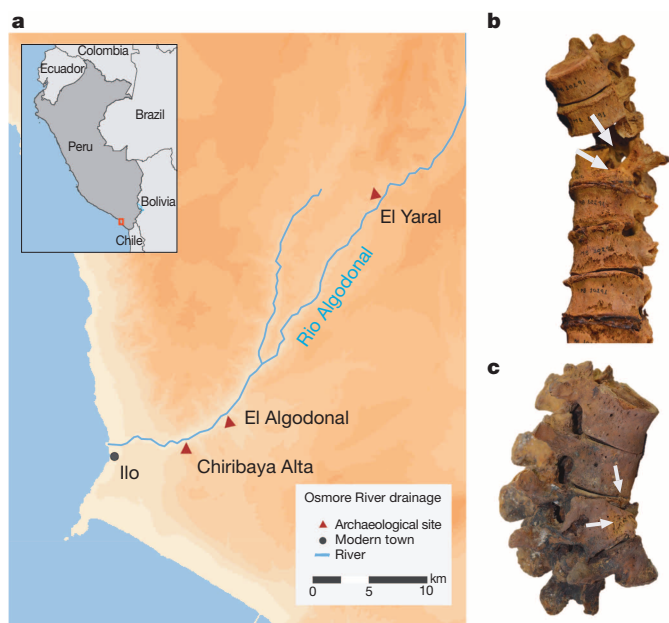
Given the pathogen's phylogeography, current models are unable to explain the abundant archaeological evidence for the presence of tuberculosis in the Americas before European contact. Strains currently circulating in the Americas are most closely related to those of European origin, and this has been used to support a European dissemination from either early settlement or trade associations<sup>1</sup>. This model, however, is incompatible with bioarchaeological data indicating the presence of tuberculosis in the pre-contact New World<sup>2</sup> (see Supplementary Information). Molecular investigations using ancient pre-Columbian material have identified short conserved regions of mobile elements considered to be diagnostic for tuberculosis, although these markers offer no information about phylogenetic placement, and are thus difficult to authenticate as ancient<sup>11</sup>. While a Pleistocene dispersal following human movements out of Africa could explain its presence in the pre-contact New World<sup>3,4</sup>, the dominance of European-derived lineages in the Americas today makes this difficult to reconcile without data to support a complete strain replacement within the past 500 years.

Genomic reconstructions of ancient pathogens provide robust evidence of DNA authenticity and permit genome-level comparisons<sup>12</sup>. The success of DNA capture<sup>13</sup> and genomic assembly of an historical MTBC strain via metagenomic sequencing<sup>14</sup> implies that DNA preservation of this pathogen may be adequate to address outstanding evolutionary questions requiring use of archaeological material. Here we apply these techniques to demonstrate that a previously uncharacterized member of the MTBC caused human infection in the Americas before European contact.

We screened 68 skeletal samples representing New World pre- and post-contact sites (Supplementary Table 1). All individuals showed skeletal indicators associated with tuberculosis infections. Samples were processed via established protocols and were screened for *M. tuberculosis* DNA by an in-solution capture assay designed for the *rpoB*, *gyrA*, *gyrB*, *katG*, and *mpt40* genes (Supplementary Table 2). Capture products for samples and negative controls were sequenced on an Illumina MiSeq and mapped to the corresponding regions in the *M. tuberculosis* H37Rv reference genome (NC\_000962.2). No tuberculosis-specific

<sup>1</sup>Department of Archaeological Sciences, University of Tübingen, Rümelinstraße 23, 72070 Tübingen, Germany. <sup>2</sup>School of Human Evolution and Social Change, Arizona State University, PO Box 872402, Tempe, Arizona 85287-2402, USA. <sup>3</sup>Center for Bioinformatics, University of Tübingen, Sand 14, 72076 Tübingen, Germany. <sup>4</sup>Department of Medical Parasitology and Infection Biology, Swiss Tropical and Public Health Institute, Socinstrasse 57, 4002 Basel, Switzerland. <sup>5</sup>University of Basel, Petersplatz 1, CH-4003 Basel, Switzerland. <sup>6</sup>Genomics and Health Unit, FISABIO-Public Health, Avenida Cataluña 21, 46020 Valencia, Spain. <sup>7</sup>CIBER (Centros de Investigación Biomédica en Red) in Epidemiology and Public Health, Instituto de Salud Carlos III, C/ Monforte de Lemos 3-5, Pabellón 11, Planta 0, 28029 Madrid, Spain. <sup>8</sup>Pathogen Genomics, The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK. <sup>9</sup>Department of Archaeology, University of Cape Town, Private Bag X1, Rondebosch, 7701, South Africa. <sup>10</sup>CONICET, Laboratorio de Ecología Evolutiva Humana (FACSO, UNCPBA), Departamento de Biología (FCEyN, UNMDP), Calle 508 No. 881 (7631), Quequen, Argentina. <sup>11</sup>Department of Anthropology, University of Tennessee, 250 South Stadium Hall, Knoxville, Tennessee 37996, USA. <sup>12</sup>Department of Anthropology, Indiana University, 701 East Kirkwood Avenue, Bloomington, Indiana 47405-7100, USA. <sup>13</sup>Molecular Mycobacteriology, Forschungszentrum Borstel, Parkallee 1, 23845 Borstel, Germany. <sup>14</sup>German Center for Infection Research, Forschungszentrum Borstel, Parkallee 1, 23845 Borstel, Germany. <sup>15</sup>McGill International TB Centre, McGill University, 1650 Cedar Avenue, Montreal H3G 1A4, Canada. <sup>16</sup>Biotechnology Institute, CICVyA-INTA Castelar, Dr. Nicolás Repetto y De Los Reseros S/N, (B1686IGC) Hurlingham, Buenos Aires, Argentina. <sup>17</sup>Instituto de Investigaciones Marinas y Costeras (CONICET-UNMDP), Facultad de Ciencias Exactas y Naturales, Universidad Nacional de Mar del Plata, San Luis 1722, Mar del Plata 7600, Argentina. <sup>18</sup>Department of Medicine, Imperial College, London W2 1PG, UK. <sup>19</sup>Division of Mycobacterial Research, MRC National Institute for Medical Research, Mill Hill, London NW7 1AA, UK. <sup>20</sup>Senckenberg Centre for Human Evolution and Palaeoenvironment, University of Tübingen, Tübingen 72070, Germany. <sup>21</sup>Max Planck Institute for Science and History, Kernalaische Straße 10, 07745 Jena, Germany.

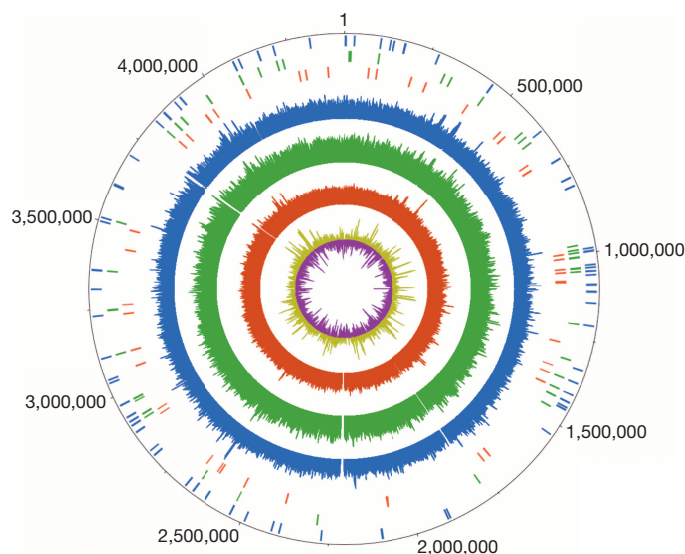
\*These authors contributed equally to this work.



**Figure 1 | Archaeological description of the skeletal samples.** **a**, Map of Peru showing locations of archaeological sites; CGIAR SRTM 90m Digital Elevation Database version 4.1 (<http://srtm.csi.cgiar.org>). **b**, **c**, Skeletal lesions of active tuberculosis from two individuals positive for *M. tuberculosis* DNA (**b**, individual 58; **c**, individual 64). Arrows show vertebral lesions, collapse, fusion, and kyphosis.

fragments were found in our negative controls. Only three of the 68 samples, referred to here as samples 54, 58, and 64, showed convincing preservation of tuberculosis DNA (see Supplementary Information, Extended Data Fig. 1, and Supplementary Table 1): all three samples were recovered from excavations in Peru and derive from Chiribaya cultures associated with the Middle Horizon/Late Intermediate period (AD 750–1350) (Fig. 1). Radiocarbon dates ranging from AD 1028 to AD 1280 (at not less than 98.5% probability) (Supplementary Table 3) confirm that they predate European contact. Spectra of DNA damage displayed a pattern expected of ancient molecules<sup>15</sup>. For comparison, non-enriched libraries were sequenced on an Illumina MiSeq producing 34,780 to 112,428 reads for each of the three samples, of which 4.6% to 1.6% mapped to the human genome (hg19). In contrast, a maximum of only 1.8% of the reads mapped to the *M. tuberculosis* reference genomes (Supplementary Table 1), indicating that DNA capture would be necessary for genome retrieval.

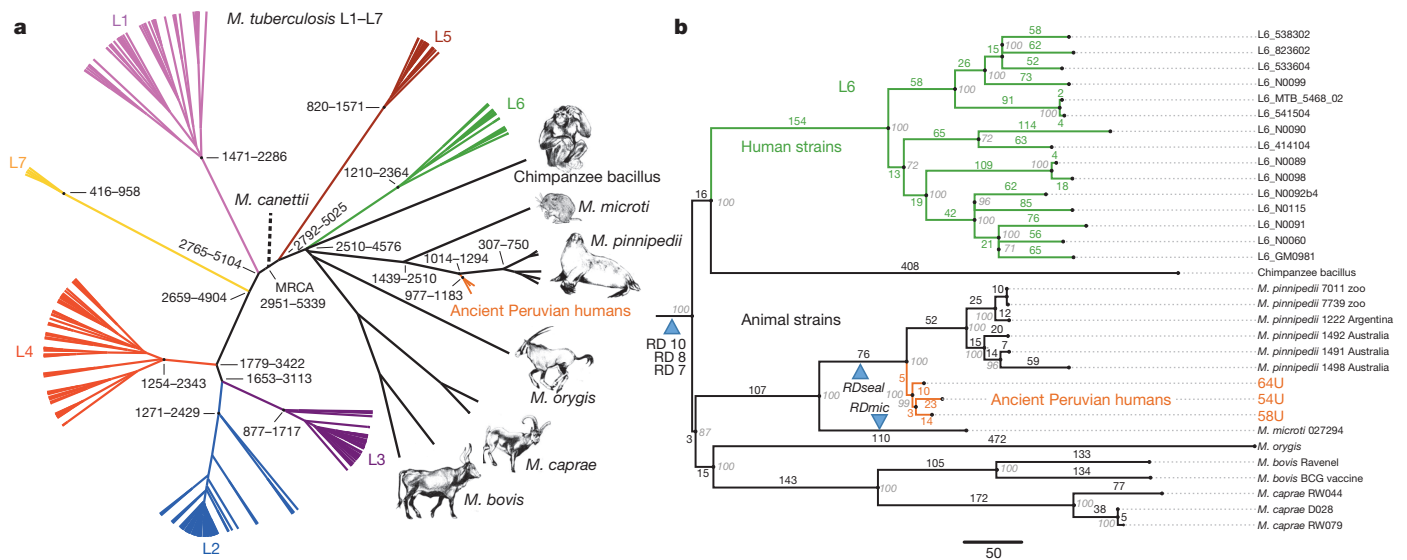
DNA libraries treated with uracil DNA glycosylase were generated to remove and repair damaged nucleotides, and were subsequently used for full genome hybridization capture (Agilent). Array probes were designed to accommodate known genetic diversity in the MTBC, as well as portions of the *Mycobacterium avium* and *Mycobacterium kansasii* genomes (Supplementary Table 4). Enriched products were sequenced on one lane of an Illumina HiSeq 2000. For comparison against a larger data set of 259 modern MTBC genomes including the outgroup *Mycobacterium canettii*<sup>3</sup>, all ancient reads were mapped against a computationally constructed ancestor for the MTBC<sup>9</sup>. The recently published genome from an eighteenth century Hungarian mummy<sup>14</sup>, as well as 14 animal strains from the *Mycobacterium caprae*, *Mycobacterium microti*, and *Mycobacterium pinnipedii* lineages, were added, along with a strain recently isolated from a wild chimpanzee<sup>16</sup>. Standard mapping resulted in heterozygous positions for the mummy, Peruvian samples 54 and 64, and all modern samples (Extended Data Fig. 2). Increased mapping stringency removed many heterozygous positions for the Peruvian samples, suggesting they derived from non-tuberculosis reads; however, the Hungarian mummy and eight modern samples still displayed heterozygosity consistent with mixed strains<sup>14</sup> (Extended Data Fig. 3). Our



**Figure 2 | Coverage plots for three ancient genomes.** Inner ring: purple, AT content; gold, GC content. Coverage rings for samples 64, 58, and 54 shown in red, green, and blue, respectively. Vertical lines indicate locations of unique SNPs. SNPs were identified before exclusion of positions with missing data from the full 262 genome data set.

more stringent mapping reduced overall genomic coverage for all samples. The final data set thus consisted of 262 genomes with a minimum of 75% coverage, four of which were ancient (Supplementary Table 5). A minimum of 20-fold average coverage was obtained for each of the Peruvian genomes (Fig. 2 and Supplementary Table 6), implying a 40- to 120-fold enrichment (Supplementary Table 6).

Single nucleotide polymorphism (SNP) analyses were performed by comparing all genomes against the constructed ancestor. This identified 53,177 SNPs for the entire data set, which ranged between 489 and 1,415 per genome (Supplementary Table 8). As input for phylogenetic assessments we used an alignment of 22,480 variable positions after removing all positions with missing data. Tree reconstructions revealed that our Peruvian genomes did not cluster with other human strains, but rather were more closely related to the animal lineage (Extended Data Figs 4–6), sharing 76 SNPs with modern *M. pinnipedii* strains (Fig. 3). Genomic architecture revealed a region of difference (RD) deletion pattern common to all animal lineages (Supplementary Table 7), as well as absence of the *M. microti*-specific *RDmic* and presence of the *M. pinnipedii*-specific *RDseal*. To our knowledge, *M. pinnipedii* strains have been isolated only from seal species restricted to the Southern Hemisphere<sup>17</sup>. Here they were harvested from captive and wild animals from South America and Australia. The three ancient strains share five unique SNPs, all of which are non-synonymous (Supplementary Table 9); this indicates that these strains derive from a common progenitor, with subsequent accumulation of 10–23 substitutions along the three strain-specific branches. To investigate possible signals of adaptation, we screened these five shared SNPs for putative functional effects. Our computational analysis predicted a functional impact of the *P44L* mutation in *Rv2258c*, encoding a methyltransferase involved in ubiquinone metabolism (Supplementary Table 9). The SNP in the *ctpA* gene at codon 62 (D62N) was not predicted to have a functional impact; however, we identified two other non-synonymous SNPs (D62G and D62E), also not predicted to have functional impacts, in the same codon of *ctpA* at different positions, each in a lineage 4 modern strain. The occurrence of homoplasies is uncommon in the MTBC, and therefore potentially indicates positive selective pressure<sup>18</sup>. A site-wise analysis of positive selection on codon 62 of *ctpA* confirmed that all three SNPs may be under diversifying selection (Supplementary Table 10). The *ctp* genes encode efflux ion pumps that are thought to prevent metal accumulation in the bacterium<sup>19</sup>, hence adaptation may relate to host metal-ion availability. This notion is supported by the existence of homoplasies in other



**Figure 3 | Phylogenetic analysis.** **a**, Bayesian maximum clade credibility tree of 261 MTBC genomes (excluding Hungarian mummy), with estimated divergence dates shown in years before present using a model of population

genes of the efflux pump family in modern MTBC strains (Supplementary Table 9).

Bayesian dating analysis used radiocarbon dates as tip calibration (Supplementary Table 3). The Hungarian mummy sample was excluded because of the presence of multiple strains. A clock test rejected the molecular clock for all 258 modern genomes ( $P = 5 \times 10^{-147}$ ) (Extended Data Fig. 7). Dating analysis using a relaxed clock model and a constant population size generated a mutation rate of  $4.6 \times 10^{-8}$  substitutions per site per year ( $3 \times 10^{-8}$  to  $6.2 \times 10^{-8}$  95% highest posterior density (HPD) interval). Bayesian skyline plots revealed constant population sizes for the animal strains and clear indications of expansions in the human-adapted lineages (Extended Data Fig. 8b). An expansion model had a negligible influence on the mutation rate, generating  $4.9 \times 10^{-8}$  substitutions per site per year ( $3.4 \times 10^{-8}$  to  $6.4 \times 10^{-8}$  95% HPD), which corresponds to 0.20 and 0.21 substitutions per genome per year for a constant and expanding population model, respectively. This rate agrees well with estimates of MTBC evolution in modern epidemiological contexts<sup>20</sup>, and is more than tenfold faster than those using human dispersals out of Africa as calibration<sup>3</sup>. Our mutation rates date the most recent common ancestor (MRCA) for the MTBC (excluding *M. canettii*) at 4,449 years before present (yr BP) (2,990–6,062 yr BP 95% HPD) and 4,064 yr BP (2,951–5,339 yr BP 95% HPD) for constant size and expansion models, respectively (Extended Data Fig. 8a). This dating was corroborated by an independent analysis using the sequences from the Hungarian mummy<sup>14</sup> sample as the only ancient calibration point. We separated the individual variants of the two mummy strains by reconstructing them onto the MTBC lineage 4 phylogeny (Extended Data Fig. 9). Lengths from the terminal branches were estimated by using the number of heterozygous variants not present in the modern strains, under the assumption that both isolates were equidistant from the root of the tree. The year of death 1797 and estimated ages for the penultimate nodes were used as priors for Bayesian phylogenetic reconstruction. Using only synonymous variants, a relaxed clock, and constant population size, we estimate the age of the MRCA of the MTBC (excluding *M. canettii*) to be 5268.5 yr BP (2689.6–8417.7 95% HPD) with a synonymous substitution rate of  $7.07 \times 10^{-8}$  ( $3.70 \times 10^{-8}$  to  $1.12 \times 10^{-7}$  95% HPD) per site per year (Extended Data Fig. 10 and Supplementary Table 11). This higher substitution rate may be due to lower selective pressure on synonymous sites.

Our results provide unequivocal evidence of human infection caused by members of the MTBC in pre-Columbian South America. Our MRCA, which is at least an order of magnitude younger than previous estimates<sup>3,4</sup>,

presented us with a challenge to explain how a mammalian pathogen could have reached human populations in the Americas about 10,000 years after inundation of the Bering land bridge<sup>21</sup>. The fact that our ancient genomes share a common ancestor with strains that are restricted to seals and sea lions<sup>17</sup> provides a plausible, if unexpected, route of entry into the New World: within the past 2,500 years pinnipeds probably contracted the disease from an African host species, carried the disease across ocean waters, and exploitation of marine mammals among coastal peoples of South America facilitated a zoonotic transfer of the bacterium within the first millennium AD. This parallels similar zoonoses of marine parasites acquired from seal consumption among archaeological coastal populations<sup>22</sup> (Supplementary Information).

Owing to the abundance of publications reporting morphological evidence of pre-Columbian tuberculosis in the region, the coasts of Peru and northern Chile have long been recognized in the archaeological literature as locations where tuberculosis first came into view in the New World<sup>2</sup>. Some have even suggested marine mammals as a potential source of the infection<sup>23</sup>. The three individuals considered here show pathological changes consistent with either pulmonary or disseminated tuberculosis, so a non-contagious infection acquired from consumption of contaminated animal products in each case cannot be ruled out. In the absence of these data, however, the five unique derived positions shared by the ancient Peruvian genomes may provide preliminary evidence of host specificity. All three genomes share a common ancestor that predates the radiocarbon age of our skeletal material by more than 100 years, and two SNPs show potential signals of adaptation. These observations could support a single zoonotic transfer from pinnipeds to humans between AD 700 and AD 1000 (Fig. 3). Subsequent host adaptation and dissemination is a compelling prospect for future work. If confirmed, this would constitute the first example of a zoonotic transfer followed by re-adaptation to the human host in the MTBC.

Such a model could explain the abundance of tuberculosis-like lesions in the region that accumulate beginning at approximately AD 700 (refs 24, 25). The later appearance of similar skeletal lesions in North America that first appear at about AD 900 is consistent with either a trans-continental spread of the pathogen via established trade routes<sup>26</sup> or a later independent introduction of tuberculosis from a different source. The lack of representation of this or any other American-specific strain in modern groups supports replacement by a European strain after contact that quickly moved through indigenous populations on account of additional adverse factors such as social marginalization, food insecurity, and potentially facilitative co-circulating infections that reached



epidemic levels, such as those recorded in northern North America during the decline of the fur trade<sup>27</sup>. Our data also indicate a subsequent introduction of *M. pinnipedii* to Australian seal colonies within the past 700 years (Fig. 3); the potential for similar zoonotic transfers, therefore, exists in Oceanian populations, although lesions suggestive of tuberculosis have not been identified in relevant skeletal material<sup>2</sup>.

*M. pinnipedii* has caused infection in several mammalian host species, including humans, in the context of zoo outbreaks<sup>28</sup>. Further sampling of animal-adapted MTBC from both modern and ancient contexts will be of great value in determining its range of potential host species and in clarifying directions of transmission. While a human transfer of the bacterium to marine mammals cannot be ruled out from our data, we consider this extremely unlikely: humans did not herd or farm seals, and close, regular contacts would be required for anthroponotic transmission, as is observed in domestic cattle<sup>29</sup>.

The above assertion of an introduction of MTBC via pinnipeds followed by human adaptation and subsequent transmission throughout the Americas can only be confirmed by comparison with additional North and South American pre-Columbian MTBC genomes from non-coastal groups, which remain elusive despite the inclusion of suitable material in our screening (Supplementary Table 1). In addition, our dating analyses are based on two independent approaches, although each relies on (effectively) a single calibration point. Mutation rate heterogeneity is documented in other clonal pathogens<sup>30</sup>, and the rejection of our molecular clock indicates that MTBC evolution is not constant among lineages. Additional calibration points from ancient MTBC lineages around the world will be essential to evaluate the legitimacy of our proposed models. Such caveats are of paramount importance considering the many investigations that report on members of the MTBC identified in skeletal samples that predate our inferred MRCA, or American material from periods that predate our proposed time of MTBC entry. Such claims could only be reconciled with what we propose here if (1) rate heterogeneity or horizontal gene transfer is obscuring our dating analysis, perhaps as a result of human population expansions which increase the availability of susceptible hosts and allow selection to operate more quickly, (2) the pathogens identified in the earlier archaeological material are in fact not members of the MTBC, but rather are ancestral forms that have since undergone replacements, or (3) certain techniques for MTBC identification in archaeological material lack specificity.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 3 March; accepted 19 June 2014.**

**Published online 20 August; corrected online 22 October 2014 (see full-text HTML version for details).**

- Hershberg, R. *et al.* High functional diversity in *Mycobacterium tuberculosis* driven by genetic drift and human demography. *PLoS Biol.* **6**, e311 (2008).
- Roberts, C. A. & Buikstra, J. E. *The Bioarchaeology of Tuberculosis. A Global View on a Reemerging Disease* 187–213 (Univ. Press of Florida, 2003).
- Comas, I. *et al.* Out-of-Africa migration and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. *Nature Genet.* **45**, 1176–1182 (2013).
- Wirth, T. *et al.* Origin, spread, and demography of the *Mycobacterium tuberculosis* complex. *PLoS Pathogens* **4**, e1000160 (2008).
- Cockburn, A. *The Evolution and Eradication of Infectious Diseases* (Johns Hopkins Press, 1963).
- Brosch, R. *et al.* A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proc. Natl Acad. Sci. USA* **99**, 3684–3689 (2002).
- Gagneux, S. & Small, P. M. Global phylogeography of *Mycobacterium tuberculosis* and implications for tuberculosis product development. *Lancet Infect. Dis.* **7**, 328–337 (2007).
- Gagneux, S. *et al.* Variable host-pathogen compatibility in *Mycobacterium tuberculosis*. *Proc. Natl Acad. Sci. USA* **103**, 2869–2873 (2006).
- Comas, I. *et al.* Human T-cell epitopes of *Mycobacterium tuberculosis* are evolutionarily hyperconserved. *Nature Genet.* **42**, 498–503 (2010).
- Pepperell, C. S. *et al.* The role of selection in shaping diversity of natural *M. tuberculosis* populations. *PLoS Pathogens* **9**, e1003543 (2013).
- Shapiro, B. & Gilbert, M. P. T. No proof that typhoid fever caused the Plague of Athens (a reply to Papagrigorakis *et al.*). *Int. J. Infect. Dis.* **10**, 334–340 (2006).

- Bos, K. I. *et al.* A draft genome of *Yersinia pestis* from victims of the Black Death. *Nature* **478**, 506–510 (2011).
- Bouwman, A. *et al.* Genotype of a historic strain of *Mycobacterium tuberculosis*. *Proc. Natl Acad. Sci. USA* **109**, 18511–18516 (2012).
- Chan, J. Z.-M. *et al.* Metagenomic analysis of tuberculosis in a mummy. *N. Engl. J. Med.* **369**, 289–290 (2013).
- Briggs, A. W. *et al.* Patterns of damage in genomic DNA sequences from a Neanderthal. *Proc. Natl Acad. Sci. USA* **104**, 14616–14621 (2007).
- Coscolla, M. *et al.* Novel *Mycobacterium tuberculosis* complex from a wild chimpanzee. *Emerg. Infect. Dis.* **19**, 969–976 (2013).
- Bastida, R. *et al.* Tuberculosis in a wild subantarctic fur seal from Argentina. *J. Wildl. Dis.* **35**, 796–798 (1999).
- Comas, I. *et al.* Whole-genome sequencing of rifampicin-resistant *Mycobacterium tuberculosis* strains identifies compensatory mutations in RNA polymerase genes. *Nature Genet.* **18**, 106–110 (2012).
- Botella, H. *et al.* Metallobiology of host-pathogen interactions: an intoxicating new insight. *Trends Microbiol.* **20**, 106–112 (2012).
- Bryant, J. M. *et al.* Inferring patient to patient transmission of *Mycobacterium tuberculosis* from whole genome sequencing data. *BMC Infect. Dis.* **13**, 110 (2013).
- Elias, S. A., Short, S. K., Nelson, C. H. & Birks, H. H. Life and times of the Bering land bridge. *Nature* **382**, 60–63 (1996).
- Patrucco, R. *et al.* Parasitological studies of coprolites of pre-Hispanic Peruvian populations. *Curr. Anthropol.* **24**, 393–394 (1983).
- Bastida, R. *et al.* La tuberculosis en grupos de cazadores recolectores de Patagonia y Tierra del Fuego: nuevas alternativas de contagio a través de la fauna silvestre. *Rev. Arg. Antropol. Biol.* **13**, 83–95 (2011).
- Salo, W. *et al.* Identification of *Mycobacterium tuberculosis* DNA in a pre-Columbian Peruvian mummy. *Proc. Natl Acad. Sci. USA* **91**, 2091–2094 (1994).
- Arriaza, R. T. *et al.* Pre-Columbian tuberculosis in northern Chile: molecular and skeletal evidence. *Am. J. Phys. Anthropol.* **98**, 37–45 (1995).
- Anawalt, P. R. Cultural contacts between Ecuador, west Mexico, and the American Southwest: clothing similarities. *Lat. Am. Antiq.* **3**, 114–129 (1992).
- Herring, D. A. & Sattenspiel, L. Social contexts, syndemics, and infectious disease in northern Aboriginal populations. *Am. J. Hum. Biol.* **19**, 190–202 (2007).
- Jurczynski, K. *et al.* Pinniped tuberculosis in Malayan tapirs (*Tapirus indicus*) and its transmission to other terrestrial mammals. *J. Zoo Wildl. Med.* **42**, 222–227 (2011).
- Berg, S. *et al.* The burden of mycobacterial disease in Ethiopian cattle: implications for public health. *PLoS ONE* **4**, e5068 (2009).
- Cui, Y. *et al.* Historical variations in mutation rate in an epidemic pathogen, *Yersinia pestis*. *Proc. Natl Acad. Sci. USA* **110**, 577–582 (2012).

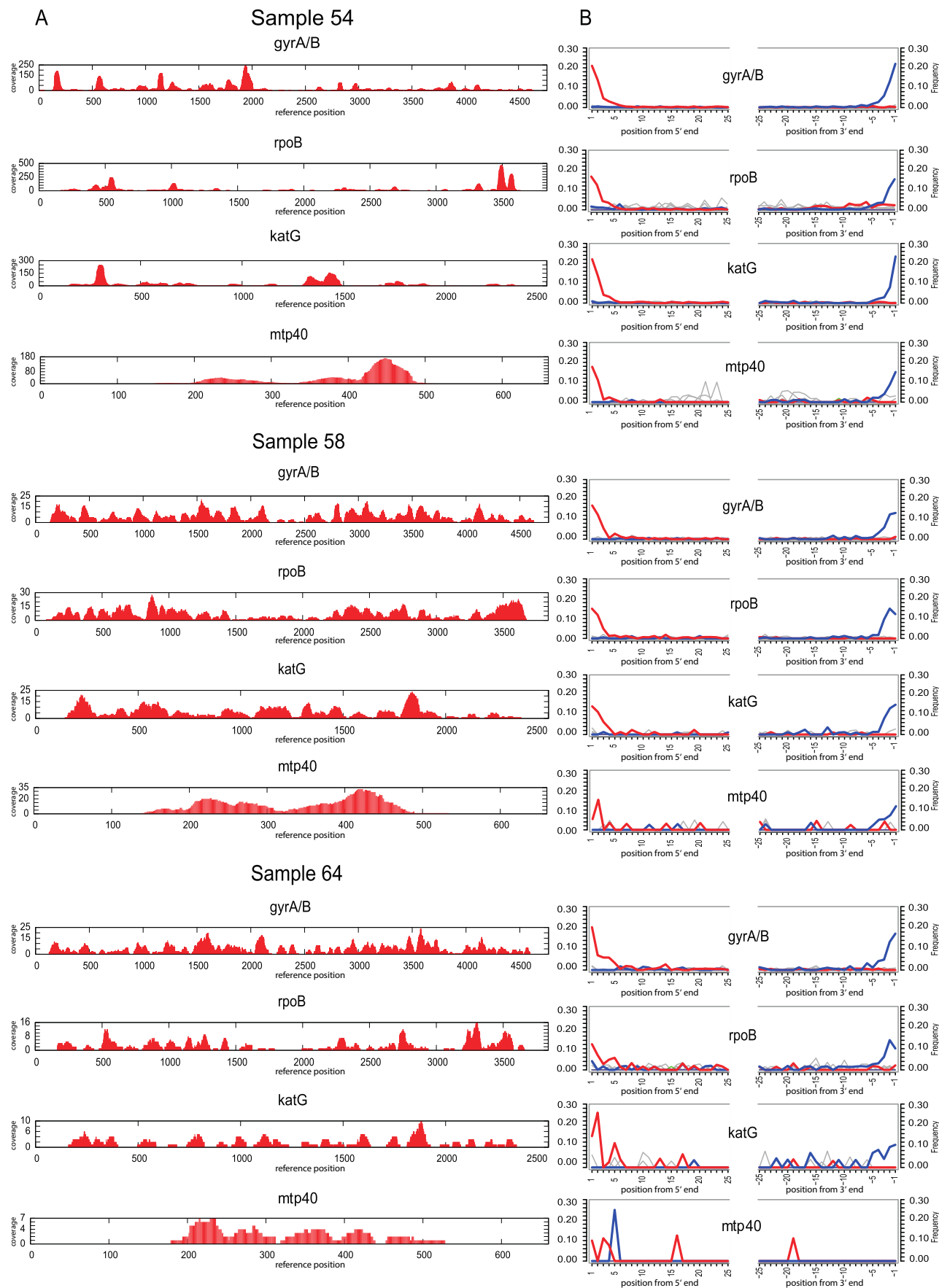
**Supplementary Information** is available in the online version of the paper.

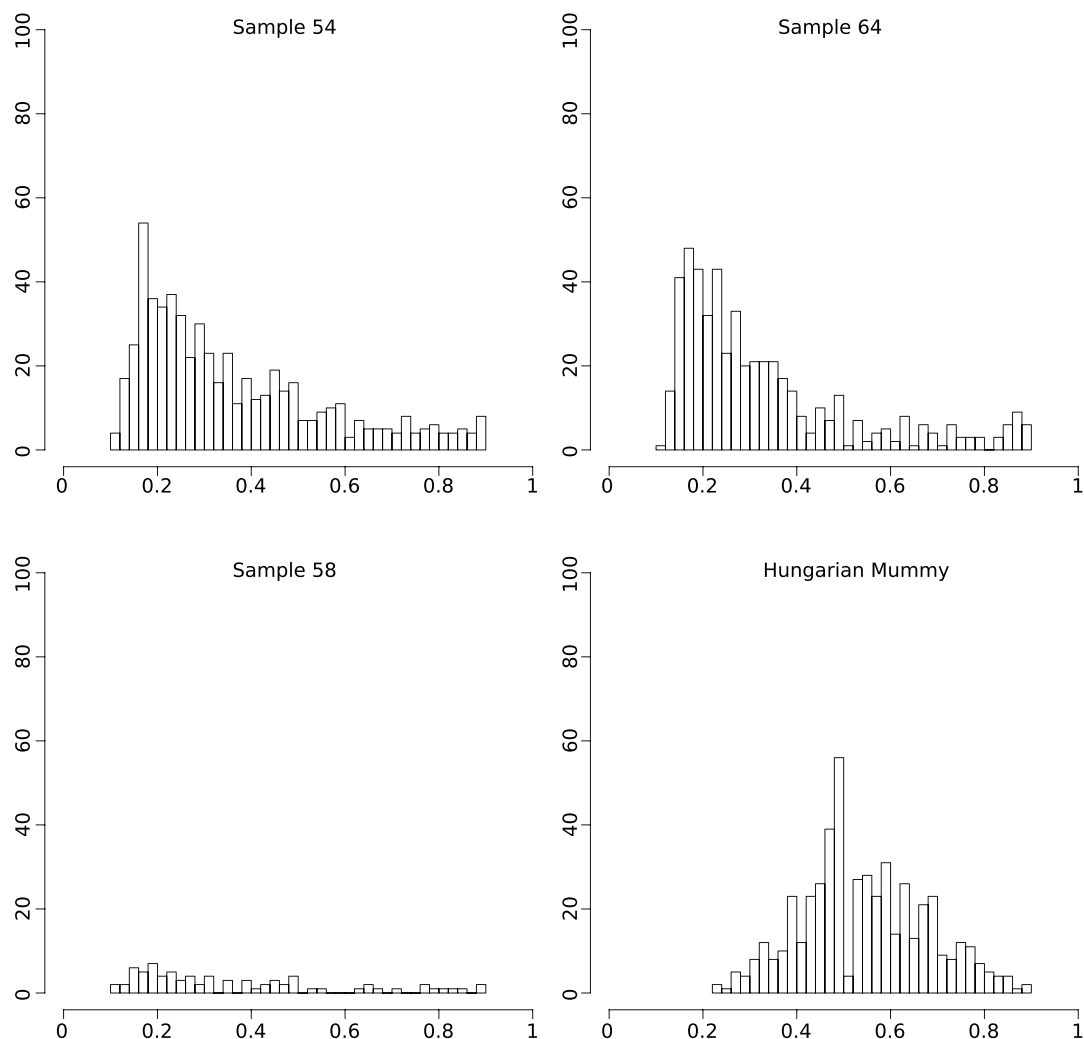
**Acknowledgements** We thank the following people and institutions for assistance and/or permission for sampling: Museo Contisuyo, Centro Mallqui, S. Guillén, Instituto Nacional de Cultura, Peru, G. Cock, C. Gaither, M. Murphy, M. C. Lozada, S. Burgess, D. Blom, B. Owen, A. Oquiche Hernani, P. Palacios Filinich, S. Williams, B. Vargas, D. Rice, and H. Klaus, S. Pfeiffer, the University of Toronto, the Upper Mississippi Valley Archaeological Research Foundation, L. Conrad, Indiana University, G. Milner, the Pennsylvania State University, the American Museum of Natural History, A. Stodder, B. Brier, I. Tattersall, K. Mowbray, the National Museum of Natural History (Smithsonian Institution), B. Billeck, the Smithsonian Institution, N. Tuross, L.-A. Pfister, Rochester Museum & Science Center, L. P. Saunders, C. Grivas, G. Housman, and M. Nieves-Colon. We thank H. Poinar for discussions about capture regions for *M. tuberculosis* screening. We thank B. Coombes and B. Krause-Kyora for providing modern tuberculosis for bait manufacture. The Huron Wendat Nation is aware of the sampling of Uxbridge bone and is a recipient of information from this study. We acknowledge the following sources of funding: European Research Council starting grant APGREID (to J.K.), the National Science Foundation (to A.C.S. and J.E.B.) for NSF BCS-1063939, NSF-REU BCS-0612222, and NSF BCS-0612222, the George E. Burch Fellow in Theoretic Medicine and Affiliated Sciences at the Smithsonian Institution (2003–2007, to J.E.B.), Social Sciences and Humanities Research Council of Canada postdoctoral fellowship grant 756-2011-501 (to K.I.B.), National Science Foundation Graduate Research Fellowship DGE-1311230 and Jacob K. Javits Fellowship (to K.M.H.), Ramón y Cajal Spanish research grant RYC-2012-10627 (to I.C.), Swiss National Science Foundation PPO033\_119205 (to S.G.), National Institutes of Health AI090928 (to S.G.), European Research Council 309540 (to S.G.), PICT0575 Argentina (to R.A.G.), Wadsworth Fellowship from the Wenner-Gren Foundation (to T.J.C.), Wellcome Trust 098051 (to J.P., J.M.B. and S.R.H.), and funding from the Medical Research Council (to J.M.B.).

**Author Contributions** A.C.S., J.E.B., J.K., K.I.B., and K.M.H. conceived the investigation. J.K., K.I.B., A.C.S., S.A.F., N.W., and A.K.W. designed experiments. J.P., R.A.G., D.L.W.S., D.C.C., S.N., M.A.B., M.Z., and R.B. provided samples for analysis. K.I.B., K.M.H., V.J.S., T.J.C., and A.K.W. performed laboratory work. A.H., J.K., S.G., M.C., N.W., K.I.B., I.C., D.Y., J.P., J.M.B., S.R.H., D.H., K.N., A.C.S., K.M.H., J.E.B., T.J.C., D.C.C., and D.L.W.S. performed analyses. K.I.B. wrote the manuscript with contributions from all co-authors.

**Author Information** Raw sequencing data have been deposited in the National Center for Biotechnology Information Sequence Read Archive under accession numbers SRP041177 for the ancient Peruvian samples and SRP041181 for the *M. pinnipedii* strains. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.I.B. ([kirsten.bos@ifu.uni-tuebingen.de](mailto:kirsten.bos@ifu.uni-tuebingen.de)), A.C.S. ([acstone@asu.edu](mailto:acstone@asu.edu)), or J.K. ([johannes.krause@uni-tuebingen.de](mailto:johannes.krause@uni-tuebingen.de)).

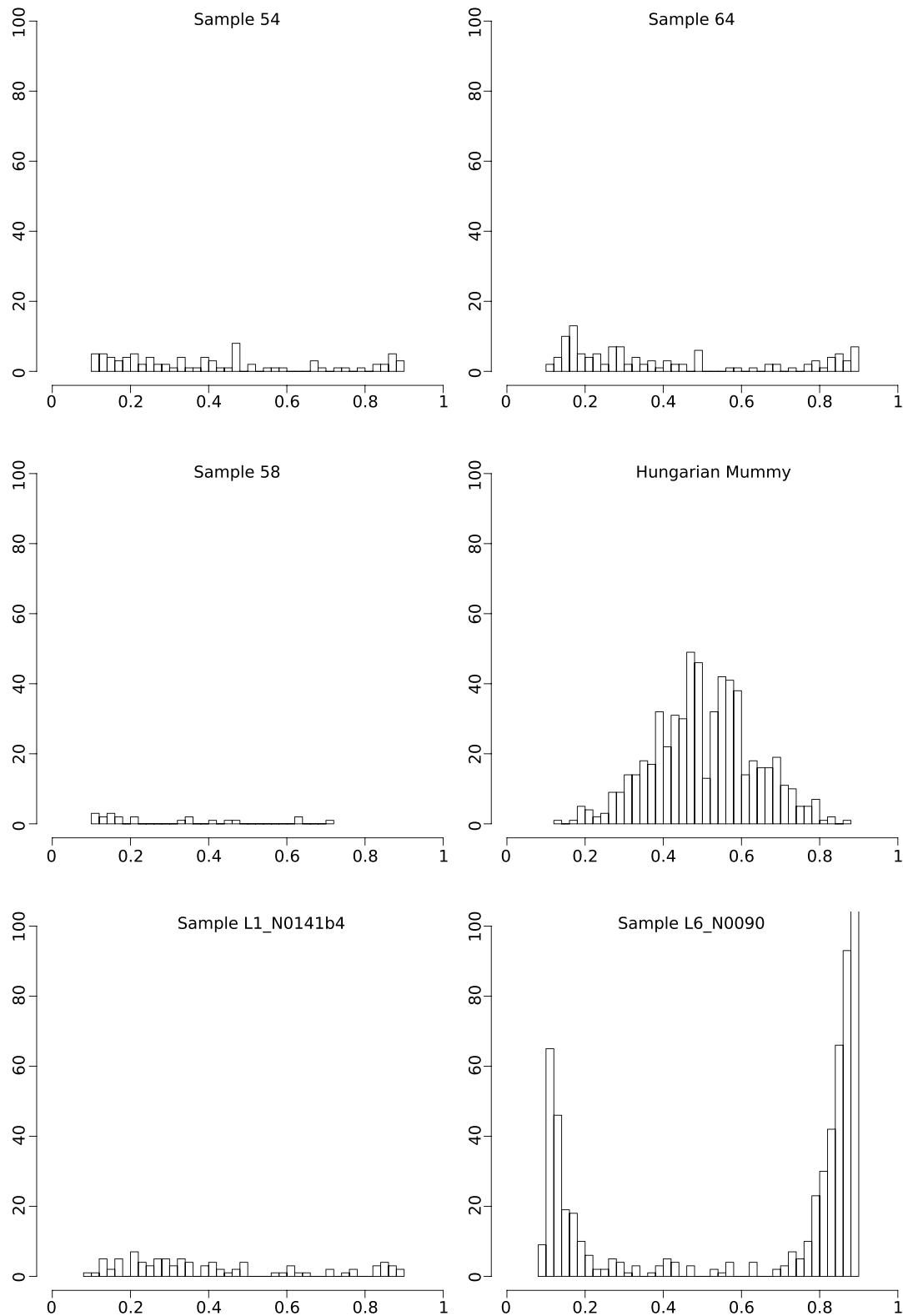


Extended Data Figure 1 | Coverage and damage plots for the *M. tuberculosis* capture regions for samples 54, 58, and 64.



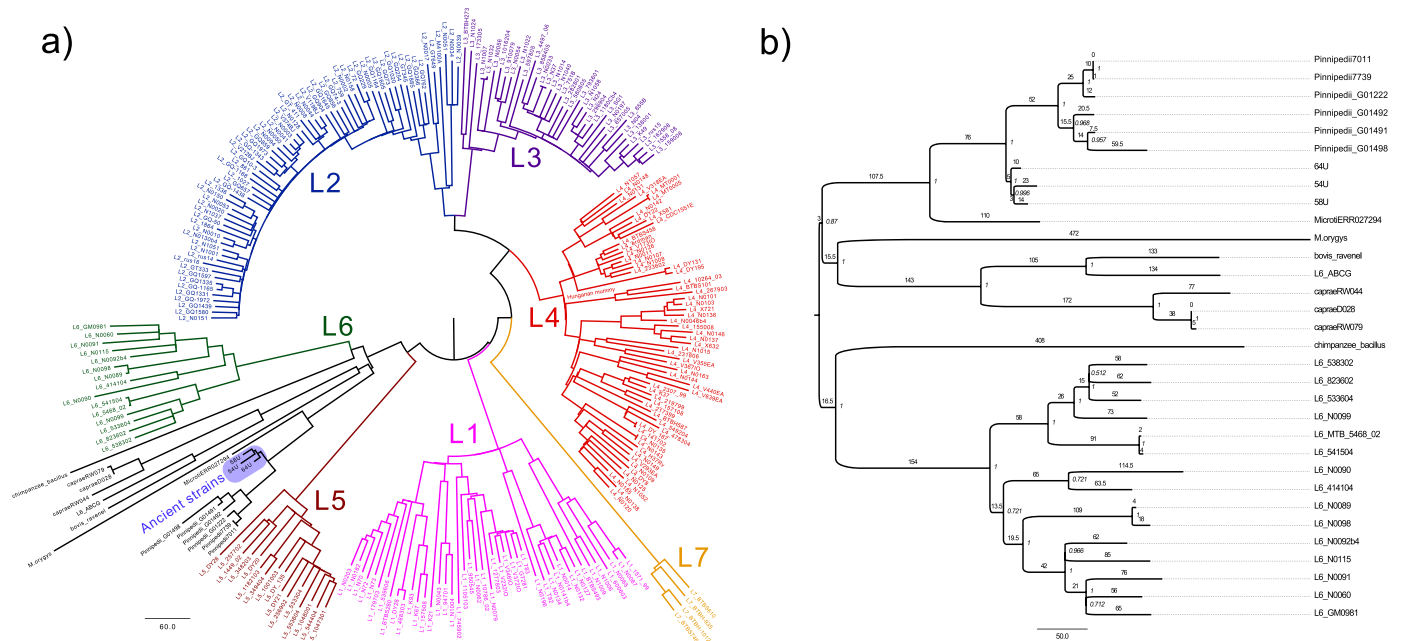
**Extended Data Figure 2 | Histograms of SNP allele frequency distributions for the ancient samples and the Hungarian mummy sample using standard mapping parameters.** The  $x$  axis denotes the frequency of reads covering a

SNP position in which the SNP was detected. The  $y$  axis denotes the number of observed SNP calls with the respective frequency. All variants with a SNP allele frequency below 90% are shown.



**Extended Data Figure 3 | Histograms of SNP allele frequency distributions for the ancient samples, the Hungarian mummy sample, and two modern isolates using stricter mapping and filtering parameters.** The  $x$  axis denotes the frequency of reads covering a SNP position in which the SNP was

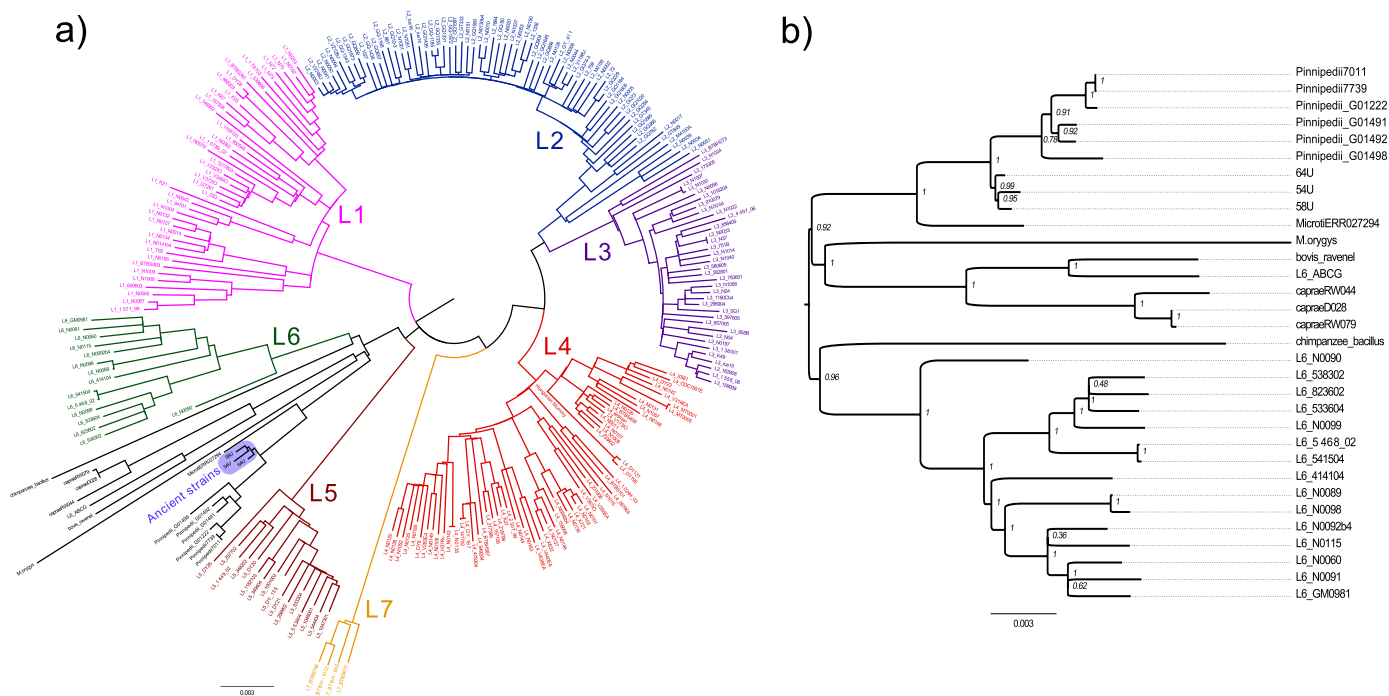
detected. The  $y$  axis denotes the number of observed SNP calls with the respective frequency. All variants with a SNP allele frequency below 90% are shown.



**Extended Data Figure 4 | Maximum parsimony analysis.** **a**, Maximum parsimony tree of all 262 samples of the complete data set. Positions with missing data were excluded. **b**, Subtree of the full maximum parsimony tree showing the lineage 6 and animal strains. Positions with missing data were

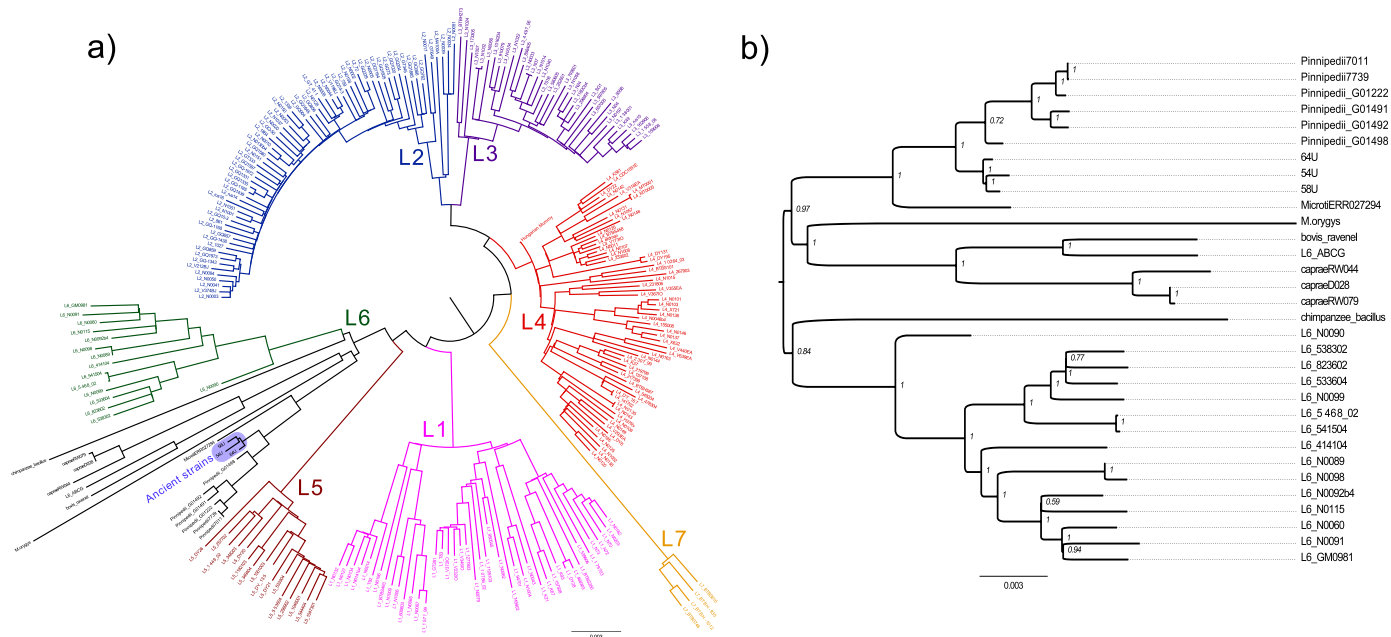
excluded. Branches are labelled with the absolute number of substitutions. Internal nodes are labelled with bootstrap statistics obtained from 1,000 replicates.



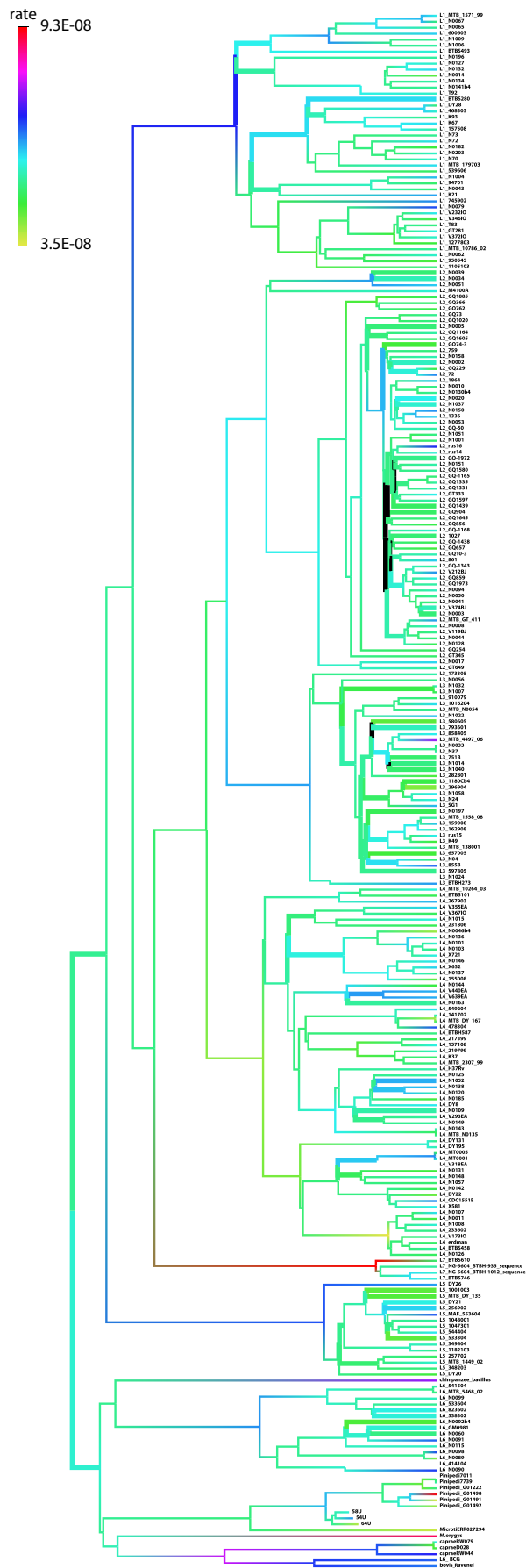


**Extended Data Figure 5 | Maximum likelihood analysis.** **a**, Maximum likelihood tree of all 262 samples of the complete data set. Positions with missing data were excluded. **b**, Maximum likelihood subtree showing the

lineage 6 and animal strains. Positions with missing data were excluded. Internal nodes are labelled with bootstrap statistics obtained from 200 replicates.

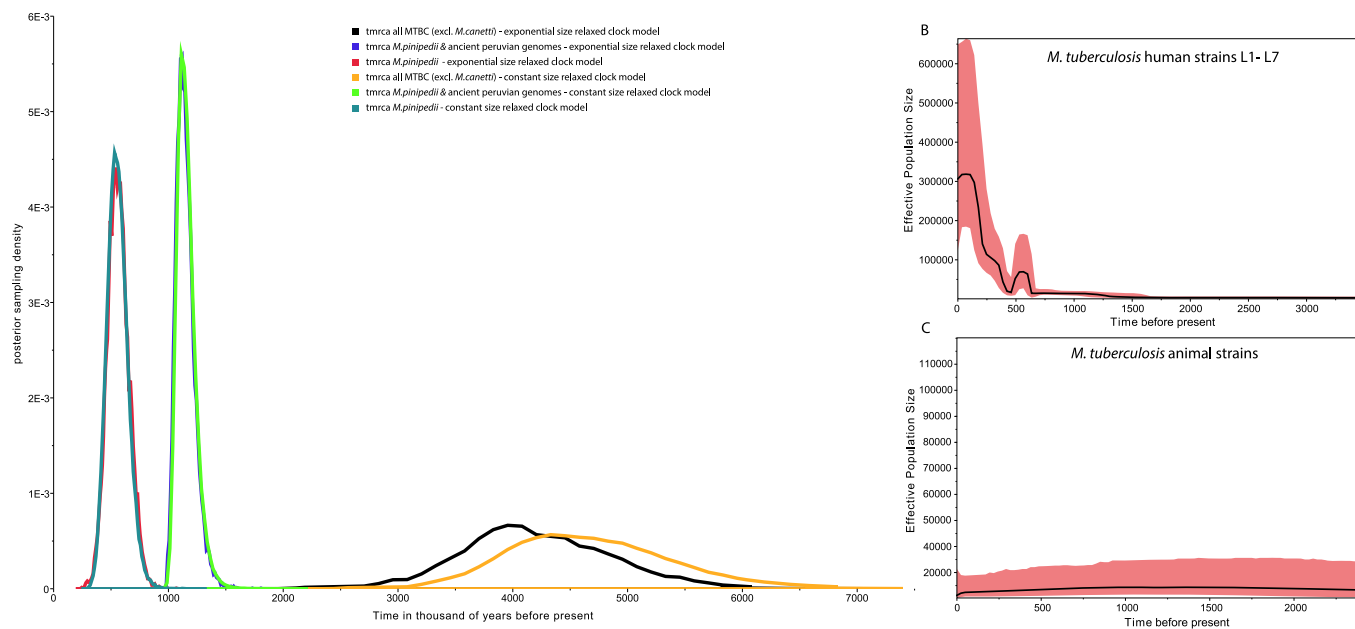


**Extended Data Figure 6 | Neighbour joining analysis.** **a**, Neighbour joining tree of all 262 samples of the complete data set. Positions with missing data were excluded. **b**, Neighbour joining subtree showing the lineage 6 and animal strains. Positions with missing data were excluded. Internal nodes are labelled with bootstrap statistics obtained from 1,000 replicates.



**Extended Data Figure 7 | Maximum clade credibility tree of *M. tuberculosis*.** The tree was estimated using the uncorrelated log-normal relaxed clock model in BEAST 1.7.5 (ref. 31). The radiocarbon dates of the ancient Peruvian strains were used as temporal estimates to date the tree. Branch lengths are scaled to years. Branch colours indicate the estimated branch substitution rate on the logarithmic scale shown in the legend at the left.

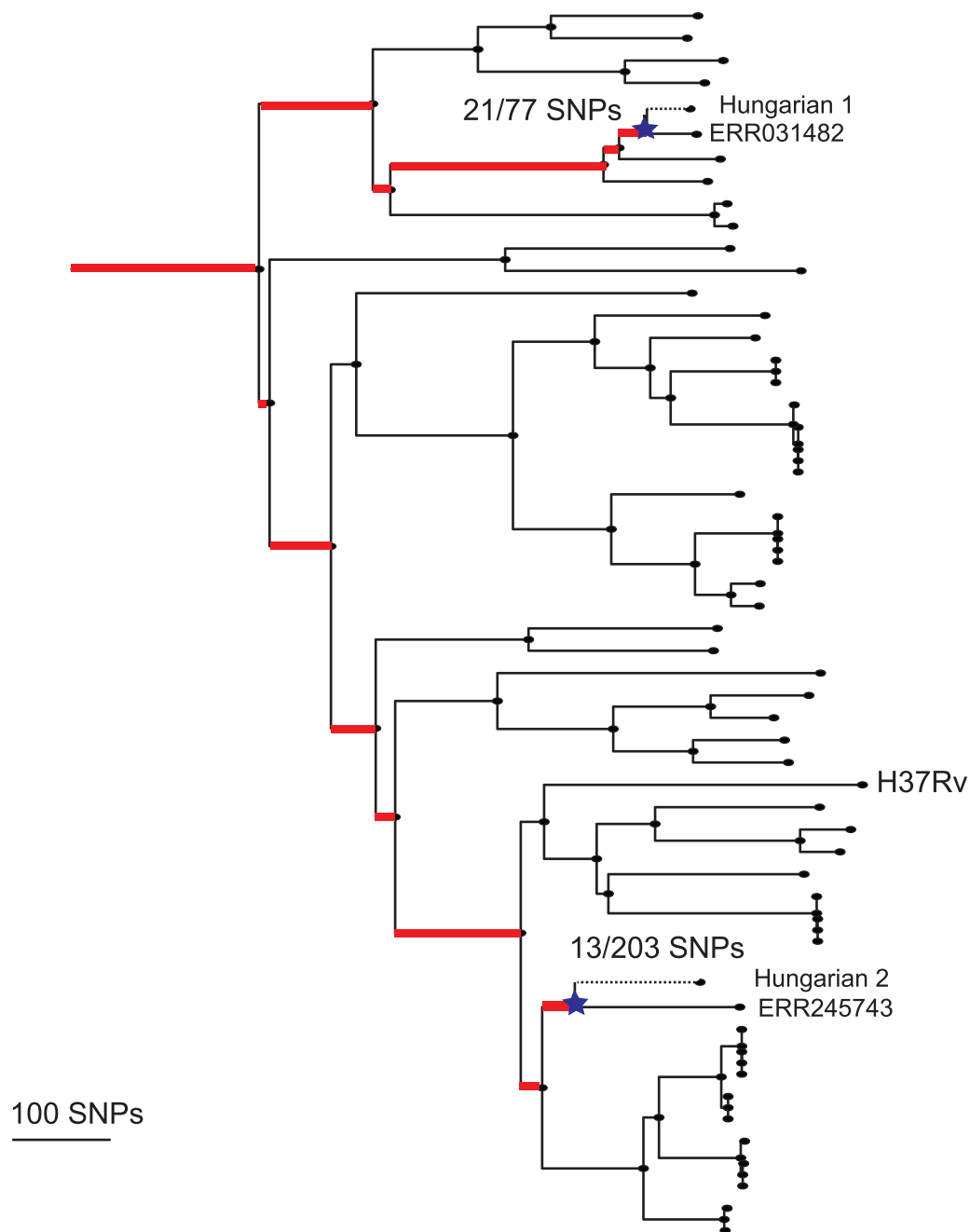
31. Drummond, A. J. & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, 214.



**Extended Data Figure 8** | **a**, Posterior distributions of times to most recent common ancestor (TMRCA) for different MTBC branches, and exponential growth and constant size models. **b**, Bayesian skyline plot showing estimated

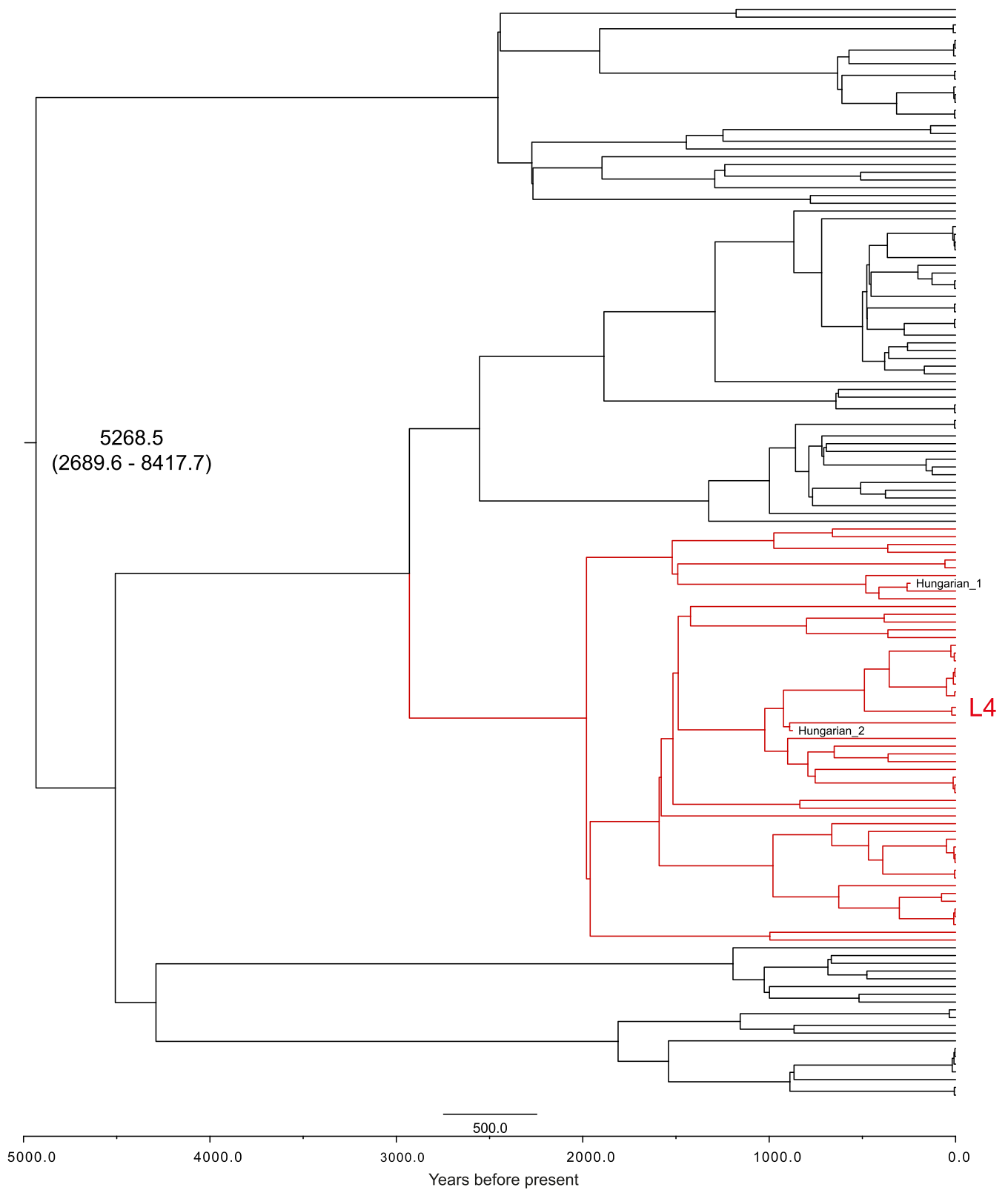
effective population sizes for the human lineages. **c**, Bayesian skyline plot showing estimated effective population sizes for the animal lineages.





**Extended Data Figure 9 | Maximum likelihood phylogeny of L4 lineage including modern and ancient strains.** The mixed samples are separated out into Hungarian 1 and 2. SNPs were mapped back onto the phylogeny, and branches marked in red are those defined by variants found to be mixed in the

Hungarian sample. This allowed us to determine the ancestral nodes and branches for each of the two strains on the tree. The dotted lines represent the unknown length of the terminal branches, with the stars representing the theoretical penultimate node for which age priors were determined.



**Extended Data Figure 10 | Maximum clade credibility tree produced using BEAST<sup>31</sup>.** Produced using TreeAnnotator from 9,000 trees. Branch lengths are scaled by age. The mean age (yr BP) of the MRCA plus 95% HPD, and the

position of the separated Hungarian ancient strains, are marked on the phylogeny.

# Dendritic cells control fibroblastic reticular network tension and lymph node expansion

Sophie E. Acton<sup>1,2</sup>, Aaron J. Farrugia<sup>1,3</sup>, Jillian L. Astarita<sup>4</sup>, Diego Mourão-Sá<sup>1,†</sup>, Robert P. Jenkins<sup>3</sup>, Emma Nye<sup>5</sup>, Steven Hooper<sup>3</sup>, Janneke van Blijswijk<sup>1</sup>, Neil C. Rogers<sup>1</sup>, Kathryn J. Snelgrove<sup>1</sup>, Ian Rosewell<sup>6</sup>, Luis F. Moita<sup>7,8</sup>, Gordon Stamp<sup>5</sup>, Shannon J. Turley<sup>9</sup>, Erik Sahai<sup>3</sup> & Caetano Reis e Sousa<sup>1</sup>

**After immunogenic challenge, infiltrating and dividing lymphocytes markedly increase lymph node cellularity, leading to organ expansion<sup>1,2</sup>. Here we report that the physical elasticity of lymph nodes is maintained in part by podoplanin (PDPN) signalling in stromal fibroblastic reticular cells (FRCs) and its modulation by CLEC-2 expressed on dendritic cells. We show in mouse cells that PDPN induces actomyosin contractility in FRCs via activation of RhoA/C and downstream Rho-associated protein kinase (ROCK). Engagement by CLEC-2 causes PDPN clustering and rapidly uncouples PDPN from RhoA/C activation, relaxing the actomyosin cytoskeleton and permitting FRC stretching. Notably, administration of CLEC-2 protein to immunized mice augments lymph node expansion. In contrast, lymph node expansion is significantly constrained in mice selectively lacking CLEC-2 expression in dendritic cells. Thus, the same dendritic cells that initiate immunity by presenting antigens to T lymphocytes<sup>3</sup> also initiate remodelling of lymph nodes by delivering CLEC-2 to FRCs. CLEC-2 modulation of PDPN signalling permits FRC network stretching and allows for the rapid lymph node expansion—driven by lymphocyte influx and proliferation—that is the critical hallmark of adaptive immunity.**

Lymph nodes are meeting places for T lymphocytes and antigen-presenting dendritic cells<sup>1,2</sup>. T-cell–dendritic-cell interactions are supported by FRCs<sup>4,5</sup>, a complex interconnected network that produces and ensheathes extracellular matrix components<sup>6</sup> that filter draining lymph<sup>7</sup>. FRC networks additionally provide physical routes for leukocyte traffic<sup>1</sup>, and chemoattractants for T cells and dendritic cells<sup>5</sup>. Furthermore, contact with FRCs promotes chemokinesis in dendritic cells, facilitating their migration within lymph nodes<sup>8</sup>. This is partly due to cytoskeletal changes in dendritic cells induced upon signalling by the C-type lectin receptor CLEC-2 when it is engaged by PDPN expressed on FRCs<sup>8</sup>. We asked whether, in addition to working as a PDPN receptor and promoting dendritic cell movement along FRCs, CLEC-2 might also act as a ligand, modulating PDPN function and altering the properties of the FRC network.

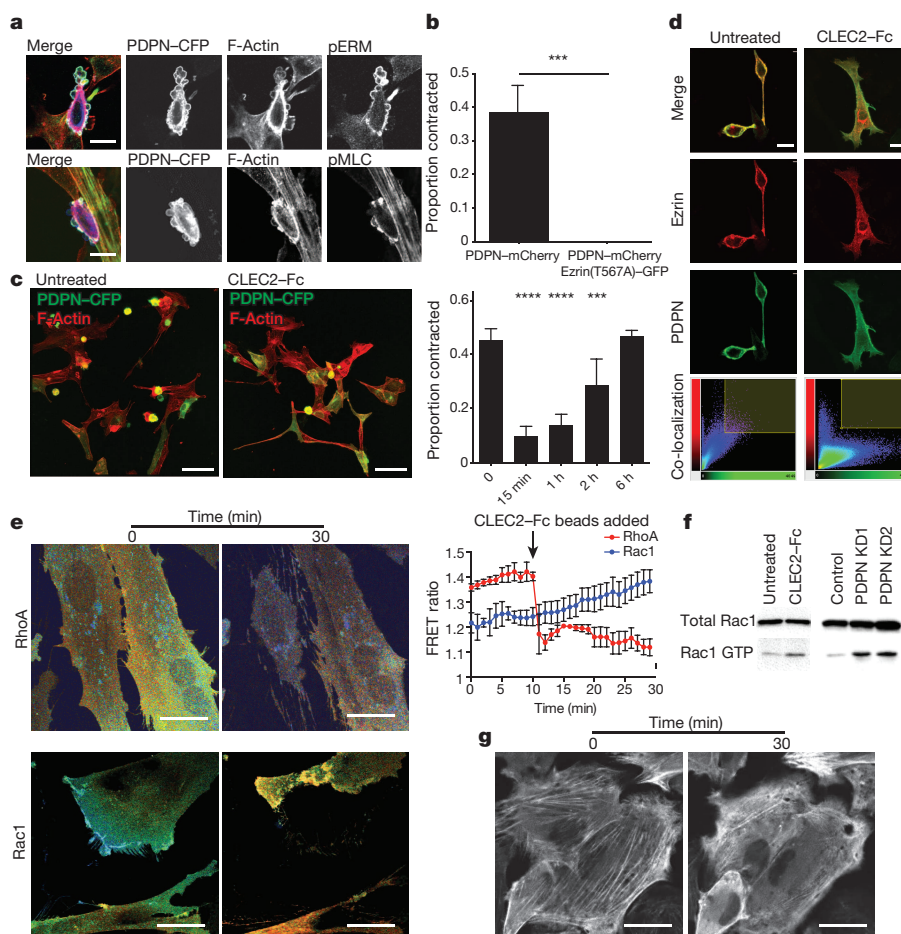
To examine PDPN signalling in fibroblasts, wild-type PDPN tagged with cyan fluorescent protein (CFP) was overexpressed in NIH/3T3 cells, which express only low levels of the endogenous protein<sup>8</sup>. Within 30 h of transfection, CFP was detectable at the plasma membrane, where it co-localized with mCherry-tagged ezrin, consistent with reports of a direct interaction between the two proteins<sup>9,10</sup> (Fig. 1a and Supplementary Video 1). Ezrin belongs to a family of closely related proteins, ezrin, radixin and moesin (ERM), which tether the actin cytoskeleton to the plasma membrane. We therefore examined the localization and phosphorylation of ERM proteins, along with myosin light chain (MLC), which mediates actin-dependent contraction, in PDPN–CFP-overexpressing

cells. In contrast to untransfected cells, PDPN–CFP<sup>+</sup> NIH/3T3 cells displayed phosphorylated (p)ERM and pMLC accumulation at the cell cortex (Fig. 1a) and often rounded up, features typical of contractile cells<sup>11–14</sup>. A non-phosphorylatable ezrin T567A mutant formally demonstrated the key role of ERM phosphorylation in PDPN-driven cell contraction (Fig. 1b).

To determine which pathways connected PDPN to cell contraction, a chemical screen was conducted, which revealed relaxation upon inhibition of RhoA/C, ROCK or myosin II family proteins (Extended Data Fig. 1a, b). Strikingly, treatment with soluble recombinant CLEC-2–Fc protein phenocopied RhoA/C and ROCK inhibition, almost completely reversing the contraction induced by PDPN–CFP (Fig. 1c). The inhibition by CLEC-2 was rapid but transient (Fig. 1c) and led to ezrin redistribution from the plasma membrane to the cytoplasm (Fig. 1d). To test this in FRCs expressing physiological levels of PDPN, we generated lymph node FRC lines (Extended Data Fig. 2 and Methods). Sublines stably expressing fluorescence resonance energy transfer (FRET) biosensors reporting RhoA or Rac1 activity were exposed to CLEC-2–Fc-coated 10 µm beads. In agreement with the NIH/3T3 studies, RhoA activity was immediately and robustly reduced when CLEC-2 beads made contact with FRCs (Fig. 1e and Supplementary Video 2). Sudden loss of RhoA activity was also evident from temporary loss of adhesion<sup>15</sup> (Supplementary Videos 2 and 3). In contrast, Rac1 activity gradually increased after exposure to CLEC-2 beads (Fig. 1e and Supplementary Video 3), which was confirmed in longer-term experiments by pull-down of GTP-bound Rac1 (Fig. 1f). Higher Rac1–GTP levels increased ARP2/3<sup>+</sup> lamellipodial protrusions, and tail retraction defects were also observed in FRCs when PDPN was stably depleted (PDPN knockdown FRCs; Fig. 1f and Extended Data Fig. 3). To identify guanine–nucleotide exchange factors (GEFs) that could connect PDPN to activation of RhoA/C, we decreased the expression of candidates using short interfering RNA (siRNA) and found that PDPN-induced contractility primarily requires GEF-H1 in NIH/3T3 and FRCs (Extended Data Fig. 1c, d). We also looked for further changes in FRCs consistent with decreased RhoA/C activity and found that PDPN knockdown or engagement by CLEC-2 caused rapid dissolution of stress fibres (Fig. 1g, Supplementary Video 4 and Extended Data Figs 3, 4). Thus, PDPN in fibroblasts associates with ezrin and signals to promote RhoA/C-dependent actomyosin-driven contraction. This is alleviated by CLEC-2 engagement, causing uncoupling of PDPN from ezrin and a RhoA/C to Rac1 switch. We hypothesize that Rac1 activity is increased indirectly as a consequence of reduced RhoA/C activity<sup>14,16</sup>.

The cytoplasmic tail of PDPN undergoes phosphorylation and ezrin recruitment requires basic residues surrounding Ser 167 (refs 10, 17). We therefore tested whether phosphorylation of Ser 167 controlled PDPN-induced contractility. Overexpression of the PDPN(S167A) mutant failed

<sup>1</sup>Immunobiology Laboratory, Cancer Research UK London Research Institute, 44 Lincoln's Inn Fields, London WC2A 3LY, UK. <sup>2</sup>Department of Cell and Developmental Biology, University College London, Gower Street, London WC1E 6BT, UK. <sup>3</sup>Tumour Cell Biology Laboratory, Cancer Research UK London Research Institute, 44 Lincoln's Inn Fields, London WC2A 3LY, UK. <sup>4</sup>Department of Cancer Immunology and AIDS, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, USA. <sup>5</sup>Experimental Histopathology Laboratory, Cancer Research UK London Research Institute, 44 Lincoln's Inn Fields, London WC2A 3LY, UK. <sup>6</sup>Transgenics Laboratory, Cancer Research UK London Research Institute, Clare Hall Laboratories, South Mimms, Potters Bar, Hertfordshire EN6 3LD, UK. <sup>7</sup>Instituto Gulbenkian de Ciência, Rua da Quinta Grande 6, 2780-156 Oeiras, Portugal. <sup>8</sup>Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, 1649-028 Lisboa, Portugal. <sup>9</sup>Department of Cancer Immunology, Genentech, One DNA Way, South San Francisco, California 94080, USA. <sup>†</sup>Present address: Laboratory of Immune Cell Epigenetics and Signaling, The Rockefeller University, 1230 York Avenue, New York, New York 10065, USA.



**Figure 1 | CLEC-2 binding uncouples PDPN from RhoA/C- and actomyosin-driven fibroblast contractility.** **a**, NIH/3T3 cells expressing PDPN-CFP (blue) or untransfected (control), fixed and stained for pERM (green) or pMLC (S19) (green) and F-actin (red). Scale bars, 20 µm. **b**, Frequency of contracting NIH/3T3 cells expressing PDPN-mCherry or PDPN-mCherry and ezrin(T567A)-GFP. **c**, NIH/3T3 cells expressing PDPN-CFP (green) stained for F-actin (red) and treated with 10 µg ml<sup>-1</sup> CLEC-2-Fc (15 min). Scale bars, 50 µm. Quantification in the right panel depicts mean  $\pm$  standard deviation (s.d.) of three experiments (>300 cells). \*\*\* $P$  < 0.0005, \*\*\*\* $P$  < 0.00005, Fisher's exact test. **d**, NIH/3T3 cells

expressing PDPN-CFP and ezrin-mCherry treated with 10 µg ml<sup>-1</sup> CLEC-2-Fc (15 min). Single optical slice (1 µm); scale bars, 20 µm. Pixel co-localization analysis is shown at the bottom. **e**, FRC cell lines expressing RhoA or Rac1 FRET biosensors exposed to CLEC-2-Fc-coated beads. Quantification of FRET ratio is shown on the right and depicts mean  $\pm$  s.d. of 15 cells from 2 independent experiments. **f**, Left: total and GTP-bound Rac1 in lysates from FRCs treated with 10 µg ml<sup>-1</sup> CLEC-2-Fc (30 min). Scale bar, 50 µm. Right: same analysis in two independent PDPN-knockdown FRC lines (KD1 and KD2) versus control line. **g**, FRC cell lines expressing green fluorescent protein (GFP)-MLC (greyscale) treated with CLEC-2-Fc-coated beads. Scale bar, 50 µm.

to cause contraction in NIH/3T3 fibroblasts and occasionally caused collapse of the cytoskeleton, potentially by inhibiting activity of low levels of endogenous PDPN (Fig. 2a). In contrast, a PDPN(S167E) phosphomimetic mutant induced contractility comparable to the wild-type protein (Fig. 2a, b). Inhibition of ROCK blocked contraction by both wild-type and S167E PDPN (Fig. 2b), placing ROCK activity downstream of Ser 167 phosphorylation. However, CLEC-2-Fc treatment, although inhibiting contraction induced by wild-type PDPN, had no effect on S167E PDPN (Fig. 2b), suggesting that regulation of the phosphorylation status of Ser 167 is one mechanism by which CLEC-2 uncouples PDPN from actomyosin contractility.

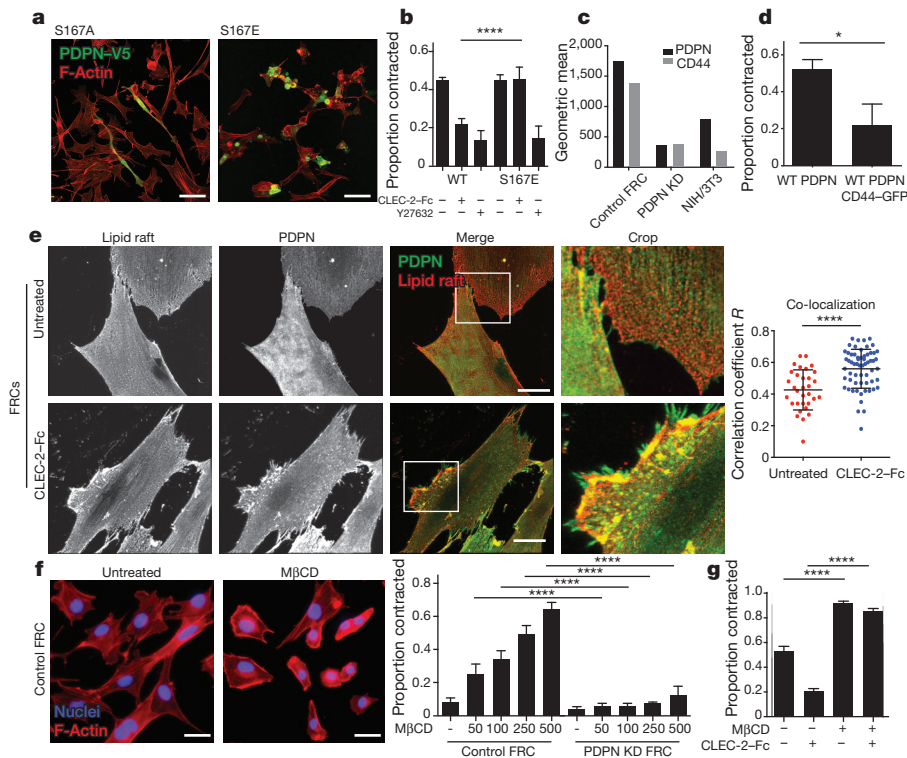
We considered how FRCs express high levels of endogenous PDPN, yet, unlike transfected NIH/3T3 cells, do not display hypercontractility. As PDPN remained at the plasma membrane when engaged with CLEC-2 (Fig. 1d), we hypothesized that it partitions between active and inactive pools, the latter being maintained by binding to an inhibitory partner such as CD44 (refs 18, 19). FRCs express high levels of both PDPN and CD44 (ref. 4) and knockdown of PDPN led to a concordant reduction of CD44 (Fig. 2c), suggesting an interaction. NIH/3T3 cells express low levels of CD44, perhaps accounting for their susceptibility to contraction after overexpression of PDPN. Consistent with that notion,

co-transfection of CD44 and PDPN into NIH/3T3 fibroblasts markedly inhibited contraction (Fig. 2d).

CD44 resides within cholesterol-rich lipid rafts<sup>19</sup> and we tested whether CLEC-2 induces PDPN redistribution to such rafts. In steady-state FRCs, PDPN and lipid rafts were found in small, partially co-localized clusters (Fig. 2e), as described in epithelial cells<sup>19</sup>. CLEC-2-Fc treatment induced formation of larger clusters in which PDPN and lipid rafts were more often co-localized (Fig. 2e). Notably, depletion of cholesterol from FRC membranes with methyl- $\beta$ -cyclodextrin (M $\beta$ CD) increased contractility, which was prevented by PDPN knockdown (Fig. 2f). CLEC-2 no longer inhibited PDPN-induced contraction in NIH/3T3 cells pre-treated with M $\beta$ CD (Fig. 2g). Together, these data support the notion that CLEC-2 sequesters PDPN within lipid rafts, where increased interaction with CD44 prevents signalling to RhoA/C. Interestingly, CD44 can itself also drive contractility when excluded from lipid rafts (data not shown), suggesting a mutually inhibitory interaction with PDPN.

To explore the biological significance of PDPN-CLEC-2 interactions for FRC function, we examined cell behaviour in three-dimensional collagen gels. Notably, FRCs reorganized the gel matrix to occupy a smaller volume, and this was inhibited by CLEC-2 treatment or PDPN knockdown (Fig. 3a). Furthermore, CLEC-2 treatment or PDPN knockdown





**Figure 2 | CLEC-2 binding causes redistribution of PDPN within the plasma membrane.** **a**, NIH/3T3 cells expressing V5-tagged PDPN mutants (PDPN-V5) stained for V5 (green) and F-actin (red). Scale bars, 50  $\mu$ m. **b**, Contraction score of NIH/3T3 cells expressing wild-type (WT) or mutant PDPN treated with 10  $\mu$ M ROCK inhibitor (Y27632) (6 h) or 10  $\mu$ g ml<sup>-1</sup> CLEC-2-Fc (30 min). Mean  $\pm$  s.d. of three independent experiments (>300 cells). \*\*\*\* $P$  < 0.00005, Fisher's exact test. **c**, Surface expression of PDPN and CD44 in the indicated cells as analysed by flow cytometry. KD, knockdown. **d**, Contraction score of NIH/3T3 cells expressing PDPN-V5  $\pm$  CD44-GFP. Mean  $\pm$  s.d. of three experiments (>150 cells). \* $P$  < 0.05, Fisher's exact test. **e**, Confocal slices (0.5  $\mu$ m) showing surface staining of lipid rafts and PDPN on

primary FRCs treated with 10  $\mu$ g ml<sup>-1</sup> CLEC-2-Fc protein (45 min). Scale bars, 50  $\mu$ m. Co-localization correlation coefficient  $R$  is shown on the right; each point represents one cell. \*\*\*\* $P$  < 0.0001, Mann-Whitney  $U$ -test. **f**, Left: FRC cell lines treated with 250  $\mu$ M M $\beta$ CD (6 h) stained for F-actin (red) and DNA (blue). Scale bars, 50  $\mu$ m. Right: contraction score in the indicated FRC cell lines treated with M $\beta$ CD. Numbers along x axis indicate dose of M $\beta$ CD in micromolar concentrations. Mean  $\pm$  s.d. of three experiments (>300 cells). \*\*\*\* $P$  < 0.00005, Fisher's exact test. **g**, Contraction score in PDPN-CFP-expressing NIH/3T3 cells pre-treated with 250  $\mu$ M M $\beta$ CD (6 h) and subsequently treated with 10  $\mu$ g ml<sup>-1</sup> CLEC-2-Fc (15 min). Mean  $\pm$  s.d. of three experiments (>300 cells). \*\*\*\* $P$  < 0.00005, Fisher's exact test.

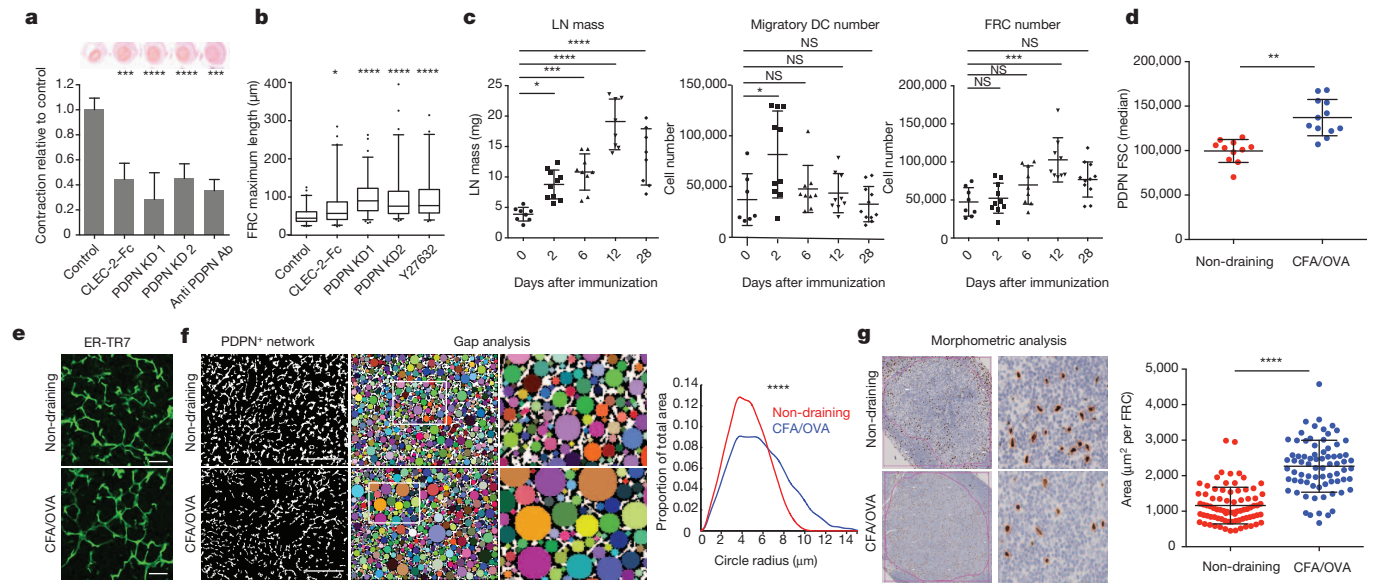
caused marked elongation of FRCs in three-dimensional culture (Fig. 3b and Extended Data Fig. 5). To determine the relevance of FRC stretching for lymph node dynamics, we investigated FRC network changes after induction of inflammation *in vivo*, which leads to upregulation of CLEC-2 by both lymph-node-resident and migratory dendritic cells<sup>8,20</sup>. We first examined the cellular composition of draining lymph nodes after subcutaneous immunization of mice with ovalbumin (OVA) in complete Freund's adjuvant (CFA). During the afferent phase (days 0–6), T- and B-cell numbers increased rapidly and total lymph node cellularity augmented 2–3 fold (Extended Data Fig. 6). However, numbers of FRCs (CD45<sup>-</sup> PDPN<sup>+</sup> CD31<sup>-</sup>) remained constant until day 6 (Fig. 3c). This lag in FRC proliferation has been previously observed<sup>21</sup>, although the kinetics probably depend on the type and strength of the inflammatory stimulus. In contrast, blood endothelial cells (CD45<sup>-</sup> PDPN<sup>-</sup> CD31<sup>+</sup>), a distinct lymph node stromal population, increased in number from the earliest time point (Extended Data Fig. 6).

If FRCs do not proliferate during the early stages of acute inflammation, then to accommodate increased lymph node size the FRC network needs to 'stretch' to avoid disruption<sup>21</sup>. Consistent with that, the mean forward scatter of lymph node FRCs, an indication of cell size, increased after immunization (Fig. 3d)<sup>21</sup> and the integrity of the FRC network was maintained (Fig. 3e). A gap analysis algorithm<sup>22</sup> revealed significantly larger spaces between the reticular network branches after immunization, consistent with FRC stretching (Fig. 3f). In a complementary approach, we used PDGFR $\alpha$ KI-H2B-GFP mice (Extended Data Fig. 7) and calculated the spacing between FRC nuclei (GFP<sup>+</sup>) using automated morphometric analysis software, which confirmed that FRCs spread further

apart after immunization (Fig. 3g). Together, these data indicate that FRCs expand and that the pre-existing FRC network enlarges in response to acute inflammation such as that following immunization.

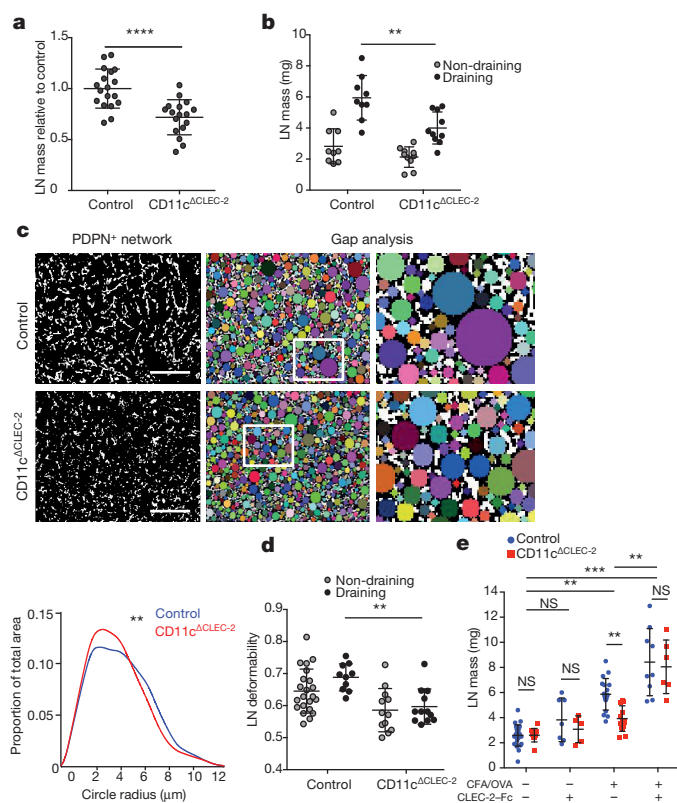
The profound cytoskeletal changes in FRCs after CLEC-2 binding *in vitro* suggested that inhibition of PDPN-induced contractility by CLEC-2<sup>+</sup> dendritic cells might aid FRC network enlargement *in vivo*. Consistent with this, lymph node influx of immigrant dendritic cells (CD45<sup>+</sup> CD11c<sup>+</sup> MHCII<sup>hi</sup>) bearing high CLEC-2 levels peaked at day 2 after OVA/CFA immunization (Fig. 3c). We therefore examined lymph node architecture and expansion in *Cd11c-cre*  $\times$  *Clec1b* <sup>$\eta/\eta$</sup>  mice (CD11c <sup>$\Delta$ CLEC-2</sup>). These mice display selective ablation of CLEC-2 in CD11c<sup>+</sup> cells (Extended Data Fig. 8), the majority of which within the T-cell zone are dendritic cells. Interestingly, in >24-week-old CD11c <sup>$\Delta$ CLEC-2</sup> mice, steady-state lymph node size was significantly reduced compared with controls (Fig. 4a), which was not the case in younger mice (Fig. 4b; non-draining lymph node). However, after immunization, expansion of draining lymph nodes was attenuated and spaces within the FRC network were smaller in young CD11c <sup>$\Delta$ CLEC-2</sup> mice compared with controls (Fig. 4b, c). Finally, the lymph nodes of CD11c <sup>$\Delta$ CLEC-2</sup> mice remained more rigid and less deformable than controls after immunization (Fig. 4d).

The attenuated expansion of lymph nodes in CD11c <sup>$\Delta$ CLEC-2</sup> mice could be due to reduced numbers of immigrating dendritic cells, decreasing antigen availability and limiting T-cell expansion. However, similar results were obtained upon administration of incomplete Freund's adjuvant without OVA or mycobacterial antigens (data not shown). Moreover, treatment by subcutaneous injection of recombinant CLEC-2-Fc protein but not a control Fc reagent (data not shown) after CFA/OVA



**Figure 3 | The FRC network stretches to accommodate acute increases in lymph node cellularity.** **a**, Collagen matrix contraction by FRC cell lines treated with  $10 \mu\text{g ml}^{-1}$  CLEC-2-Fc, anti-PDPN antibody (Ab) or stably depleted of PDPN (PDPN KD1 and 2). Representative image is shown at the top. Mean  $\pm$  s.d. of eight replicates from two experiments. \*\*\* $P < 0.0001$ , \*\*\*\* $P < 0.00001$ , one-way analysis of variance (ANOVA), Dunnett's multiple comparisons. **b**, Maximum three-dimensional length from cells as in **a** calculated from  $100 \mu\text{m}$  confocal z-stacks. Graph indicates median, 25th and 75th percentiles (range 10th–90th percentile). Data are from two independent experiments ( $>80$  cells). \* $P < 0.05$ , \*\*\*\* $P < 0.00001$ , one-way ANOVA, Dunnett's multiple comparisons. **c**, Lymph node (LN) mass, number of migratory dendritic cells (DCs;  $\text{CD}45^{+} \text{CD}11\text{c}^{+} \text{MHCII}^{\text{hi}}$ ) and number of FRCs ( $\text{CD}45^{-} \text{PDPN}^{+} \text{CD}31^{-}$ ) in lymph nodes draining the site of OVA/CFA immunization. Each point represents one lymph node. Data show mean  $\pm$  s.d. of two experiments with 8–12 mice per time point. Day 0 represents non-immunized mice. \* $P < 0.05$ , \*\* $P > 0.001$ , \*\*\* $P > 0.0001$ , \*\*\*\* $P < 0.00001$ ,

one-way ANOVA, Tukey's multiple comparisons. NS, not significant. **d**, Forward scatter (FSC) of FRCs after OVA/CFA immunization (day 6). \*\* $P < 0.001$ , Mann–Whitney  $U$ -test. **e**, ER-TR7 monoclonal antibody staining (green) of the T-cell zone of draining or non-draining lymph nodes after OVA/CFA immunization (day 6). Scale bars,  $30 \mu\text{m}$ . **f**, Left: PDPN staining (white) of lymph node sections as in **e** converted to binary images for gap analysis. Scale bars,  $100 \mu\text{m}$ . Right: gaps (coloured circles) within the FRC network. White box indicates area at higher magnification. Quantification is shown on the far right. \*\*\*\* $P = 4.194 \times 10^{-9}$  (proportion of radii  $>15 \mu\text{m}$ ), Fisher's exact test. **g**, Morphometric analysis of GFP $^{+}$  nuclei (FRCs) in draining or non-draining lymph node sections from PDGFR $\alpha$ KI-H2B-GFP mice after CFA/OVA immunization (day 6) (left). Magnification,  $\times 10$ . Anti-GFP (brown), haematoxylin (blue). Quantification of average area occupied by individual FRCs is shown on the right. \*\*\*\* $P < 0.00001$ , Mann–Whitney  $U$ -test. Data in **f** and **g** are from multiple sections from eight mice in two independent experiments.



immunization markedly augmented lymph node expansion (Fig. 4e) and, importantly, completely rescued defects in lymph node expansion in  $\text{CD}11\text{c}^{\Delta\text{CLEC-2}}$  mice (Fig. 4e). The restoration of lymph node expansion indicates that diminished CLEC-2 provision rather than a dearth of immigrant antigen-presenting dendritic cells limits such expansion in  $\text{CD}11\text{c}^{\Delta\text{CLEC-2}}$  mice. Together, these data indicate that CLEC-2 delivery by dendritic cells is required for maintaining FRC network architecture *in vivo* and is permissive for acute increases in lymph node size driven by cell influx provoked by local inflammation.

Lymph nodes are dynamic structures that must rapidly expand to accommodate leukocyte recruitment and proliferation. Given that lymph

**Figure 4 | CLEC-2 $^{+}$  dendritic cells are required for lymph node swelling during adaptive immune responses.** **a**, Mass of skin draining lymph nodes from  $\text{CD}11\text{c}^{\Delta\text{CLEC-2}}$  mice and  $\text{Cre}^{\text{neg}}$  littermate controls  $>24$  weeks old. Data are normalized to average lymph node (LN) mass of  $\text{Cre}^{\text{neg}}$  control. Each data point represents one lymph node. \*\*\*\* $P < 0.00001$ , Mann–Whitney  $U$ -test. **b**, Lymph node mass (mg) of draining and non-draining lymph nodes from 8–12-week-old  $\text{CD}11\text{c}^{\Delta\text{CLEC-2}}$  mice and  $\text{Cre}^{\text{neg}}$  littermates (control) immunized with OVA/CFA (day 7). Each point represents a lymph node. Data are from three experiments. \*\* $P < 0.001$ , two-way ANOVA, Tukey's multiple comparisons. **c**, Gap analysis of draining lymph nodes from immunized control (top) and  $\text{CD}11\text{c}^{\Delta\text{CLEC-2}}$  mice (bottom), quantified on the bottom left. \*\* $P = 0.0001759$  (radii  $>8 \mu\text{m}$ ), Fisher's exact test. Scale bar,  $100 \mu\text{m}$ . **d**, Lymph node deformation after compression of whole lymph nodes with  $1.4 \text{ N}$  force. Each point represents one lymph node. \*\* $P < 0.05$ , one-way ANOVA. **e**, Lymph node mass of  $\text{CD}11\text{c}^{\Delta\text{CLEC-2}}$  mice and  $\text{Cre}^{\text{neg}}$  littermates (control) 7 days after CFA/OVA immunization and treatment with  $10 \mu\text{g}$  CLEC-2-Fc or PBS (days 1 and 3). Each point represents one lymph node; data are from three experiments. \*\* $P < 0.001$ , \*\*\* $P < 0.0001$ , two-way ANOVA, Tukey's multiple comparisons test. NS, not significant.

nodes can expand tenfold during adaptive immune responses while maintaining integrity, stromal components must by necessity proliferate<sup>21,23</sup>. However, in the early phases of adaptive immune responses or in response to acute inflammation, FRC proliferation is insufficient to account for lymph node expansion<sup>21</sup> (Fig. 3b). We show that early lymph node expansion is permitted by FRC network relaxation, induced by increased availability of CLEC-2<sup>hi</sup> dendritic cells. We previously reported that CLEC-2 engagement by PDPN promotes dendritic cell migration along the FRC network<sup>8</sup>. Our data now show that PDPN is not simply a ligand for CLEC-2 but that it also reverse signals into FRCs to control actomyosin contractility. Our results suggest that a function of endogenous PDPN on FRCs is to cause stromal network contraction and create physical tension in lymph nodes. This is offset by constant contact between FRCs and CLEC2<sup>+</sup> resident dendritic cells, balancing contractility and controlling lymph node cellularity. The influx of additional CLEC-2<sup>hi</sup> dendritic cells, combined with the upregulation of CLEC-2 on resident dendritic cells during acute inflammation<sup>8,20</sup>, increases inhibition of PDPN, allowing the FRC network to stretch. We predict that this same mechanism can promote lymph node reduction as the short wave of increased CLEC-2 availability ends and PDPN activity returns. Whether the underlying collagen network is similarly elastic and to what extent it acts to limit or promote lymph node expansion is as yet unaddressed question. Interestingly, CLEC-2 may not be sufficient for sustained lymph node expansion, as administration of CLEC-2-Fc in the absence of inflammation did not affect lymph node size. Rather, CLEC-2 inhibition of PDPN is permissive for stretching and the influx of leukocytes via the high endothelial venules and afferent lymph probably provides the force for expansion. It is interesting to speculate that the stretching mechanism described in this study may also help initiate subsequent FRC proliferation, given the emerging connections between tension and cell cycle regulation<sup>24</sup>.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 10 March; accepted 1 September 2014.**

- Girard, J.-P., Moussion, C. & Förster, R. HEVs, lymphatics and homeostatic immune cell trafficking in lymph nodes. *Nature Rev. Immunol.* **12**, 762–773 (2012).
- Junt, T., Scandella, E. & Ludewig, B. Form follows function: lymphoid tissue microarchitecture in antimicrobial immune defence. *Nature Rev. Immunol.* **8**, 764–775 (2008).
- Banchereau, J. & Steinman, R. M. Dendritic cells and the control of immunity. *Nature* **392**, 245–252 (1998).
- Link, A. *et al.* Fibroblastic reticular cells in lymph nodes regulate the homeostasis of naive T cells. *Nature Immunol.* **8**, 1255–1265 (2007).
- Katakai, T. A novel reticular stromal structure in lymph node cortex: an immuno-platform for interactions among dendritic cells, T cells and B cells. *Int. Immunol.* **16**, 1133–1142 (2004).
- Sixt, M. *et al.* The conduit system transports soluble antigens from the afferent lymph to resident dendritic cells in the T cell area of the lymph node. *Immunity* **22**, 19–29 (2005).
- Gretz, J. E., Anderson, A. O. & Shaw, S. Cords, channels, corridors and conduits: critical architectural elements facilitating cell interactions in the lymph node cortex. *Immunol. Rev.* **156**, 11–24 (1997).
- Acton, S. E. *et al.* Podoplanin-rich stromal networks induce dendritic cell motility via activation of the C-type lectin receptor CLEC-2. *Immunity* **37**, 276–289 (2012).
- Ivetic, A. & Ridley, A. J. Ezrin/radixin/moesin proteins and Rho GTPase signalling in leucocytes. *Immunology* **112**, 165–176 (2004).
- Martin-Villar, E. *et al.* Podoplanin binds ERM proteins to activate RhoA and promote epithelial-mesenchymal transition. *J. Cell Sci.* **119**, 4541–4553 (2006).
- Olson, M. F. & Sahai, E. The actin cytoskeleton in cancer cell motility. *Clin. Exp. Metastasis* **26**, 273–287 (2009).
- Wyckoff, J. B., Pinner, S. E., Gschmeissner, S., Condeelis, J. S. & Sahai, E. ROCK- and myosin-dependent matrix deformation enables protease-independent tumor-cell invasion *in vivo*. *Curr. Biol.* **16**, 1515–1523 (2006).
- Lee, J.-W. *et al.* Peripheral antigen display by lymph node stroma promotes T cell tolerance to intestinal self. *Nature Immunol.* **8**, 181–190 (2007).
- Sanz-Moreno, V. *et al.* Rac activation and inactivation control plasticity of tumor cell movement. *Cell* **135**, 510–523 (2008).
- Wolfenson, H., Bershadsky, A., Henis, Y. I. & Geiger, B. Actomyosin-generated tension controls the molecular kinetics of cell adhesions. *J. Cell Sci.* **124**, 1425–1432 (2011).
- Nimnual, A. S., Taylor, L. J. & Bar-Sagi, D. Redox-dependent downregulation of Rho by Rac. *Nature Cell Biol.* **5**, 236–241 (2003).
- Krishnan, H. *et al.* Serines in the intracellular tail of podoplanin (PDPN) regulate cell motility. *J. Biol. Chem.* **288**, 12215–12221 (2013).
- Martin-Villar, E. *et al.* Podoplanin associates with CD44 to promote directional cell migration. *Mol. Biol. Cell* **21**, 4387–4399 (2010).
- Fernández Muñoz, B. *et al.* The transmembrane domain of podoplanin is required for its association with lipid rafts and the induction of epithelial-mesenchymal transition. *Int. J. Biochem. Cell Biol.* **43**, 886–896 (2011).
- Mourão-Sá, D. *et al.* CLEC-2 signaling via Syk in myeloid cells can regulate inflammatory responses. *Eur. J. Immunol.* **41**, 3040–3053 (2011).
- Yang, C.-Y. *et al.* Trapping of naive lymphocytes triggers rapid growth and remodeling of the fibroblast network in reactive murine lymph nodes. *Proc. Natl Acad. Sci. USA* **111**, E109–E118 (2014).
- Tozluoglu, M. *et al.* Matrix geometry determines optimal cancer cell migration strategy and modulates response to interventions. *Nature Cell Biol.* **15**, 751–762 (2013).
- Chyou, S. *et al.* Coordinated regulation of lymph node vascular-stromal growth first by CD11c<sup>+</sup> cells and then by T and B cells. *J. Immunol.* **187**, 5558–5567 (2011).
- Mammoto, A., Mammoto, T. & Ingber, D. E. Mechanosensitive mechanisms in transcriptional regulation. *J. Cell Sci.* **125**, 3061–3073 (2012).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank B. Reizis for CD11c-Cre mice, the LRI Biological Resources staff for animal care and assistance with experiments and R. Horton-Harpin for production of CLEC-2-Fc. We are grateful to M. Y. Gerner and R. N. Germain for critical reading of the manuscript and to all the members of the Immunobiology Laboratory and Tumour Cell Biology Laboratory for helpful discussions and comments. This work was supported by a Henry Wellcome Postdoctoral fellowship (S.E.A.) and core funding from Cancer Research UK (I.R., G.S., E.S. and C.R.S.).

**Author Contributions** S.E.A., E.S. and C.R.S. designed the study, analysed data and wrote the manuscript. S.E.A. conducted experiments with assistance from A.J.F., J.v.B., R.P.J., K.J.S. and S.H. Preliminary results were generated with S.J.T. and J.L.A. E.N. and G.S. carried out immunohistochemistry and morphometric analysis. D.M.-S., N.C.R. and I.R. generated CD11c<sup>ACLEC-2</sup> mice. N.C.R. managed breeding and genotyping of mouse strains. L.F.M. provided key reagents.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.E.A. ([sophie.acton@cancer.org.uk](mailto:sophie.acton@cancer.org.uk)) or C.R.S. ([caetano@cancer.org.uk](mailto:caetano@cancer.org.uk)).



## METHODS

**Mice.** Experiments were performed in accordance with national and institutional guidelines for animal care and approved by the Institutional Animal Ethics Committee Review Board, Cancer Research UK and the UK Home Office. Wild-type C57BL/6J mice were purchased from Charles River Laboratories. PDGFR $\alpha$ KI-H2B-GFP mice (B6.129S4-Pdgfratm11(EGFP)Sor/J) were purchased from Jackson Laboratories. To generate CLEC-2 floxed mice on a C57BL/6 background, C57BL/6 mouse *Clec1b* genomic regions were cloned into a pFloxRI+TK targeting vector from the bacterial artificial chromosome (BAC) clone R248K14 using the Red/Et recombination of Quick and Easy BAC modification kit (Gene Bridges). Primers used were: *Clec1b* BAC Fwd, TATTACCTGATGCTGTTACATCTCAGCTCTGCAG TATTAGCCACCTTAGAGTTCTAGCTGCTGACTCTgggtaccgagctgaattct accg; *Clec1b* BAC Rev, CTGGGTCTTTCCAGCTTCTGGCTATTATAAATAA GGCTGTTATAGCAATAGTGGAGCATGTGTCCTTCTTggggccgaccgcgggtgg agctcca. Uppercase letters indicate regions homologous to *Clec1b* regions and lowercase letters indicate regions homologous to pFloxRI+TK vector. PCR was performed using Pwo DNA polymerase (Roche) following the manufacturer's instructions.

The 5' *loxP* site was introduced in the intronic region between exon 1 and exon 2 by insertion of the *loxP*-pgk-*gb2*-NEO-*loxP* cassette (Gene Bridges) using the following primers: 1st *loxP* Fwd, AAAACCCAAAACCAAAAACCAAAAACCAAC AAAAAACAAAAACAGATaattacacctactaaaggcg; 1st *loxP* Rev, ACTTAT TCTCTGTCCATTCTAACATATAACTGGCTACCAAGGCCACGTGtaatacag actcactataggctc. Uppercase letters indicate regions homologous to *Clec1b* regions and lowercase letters indicate regions homologous to *loxP*-pgk-*gb2*-NEO-*loxP* cassette. PCR was performed using Pwo DNA polymerase (Roche) following the manufacturer's instructions.

The vector was then transformed into Cre-expressing *Escherichia coli* EL350 (gift from A. Behrens), leading to recombination of the cassette and leaving a single *loxP* site.

Next, the FRT-pgk-*gb2*-NEO-FRT-*loxP* cassette was introduced into the intronic region between exon 4 and exon 5 using the following primers: *Rb2* Fwd, tccatgtc aagcattttggaatgctgagggaaacattgaatgctgttaataacctactaaaggcg; *Rb2* Rev, tctcagag gagcacacatgcaaacattgaacacatgaaaggaaataatcagactcactataggctc. Uppercase letters indicate regions homologous to *Clec1b* regions and lowercase letters represent regions homologous to FRT-pgk-*gb2*-NEO-FRT-*loxP* cassette. PCR was performed using Pwo DNA polymerase (Roche) following the manufacturer's instructions.

The targeting vector was linearized by SfiI digestion (New England Biolabs), precipitated by phenol/chloroform and electroporated into PRX-B6N C57BL/6N embryonic stem (ES) cells.

Screening for homologous recombination in ES cells and mice was done by PCR using primers inside the FRT-pgk-*gb2*-NEO-FRT-*loxP* and outside the homology arm: *Clec1b* Gen Rev, agacctgagaagctgga; NEO 3' sense, gctccgattgcagcgcac.

Deletion of the NEO cassette was performed by crossing *Clec1b*-targeted mice with  $\beta$ -actin-Flp (B6.Cg-Tg(ACF1P)9205Dym), and thereafter generating *Clec1b*<sup>fl/fl</sup> by interbreeding and screening for deletion of the NEO cassette using the following primers: Seq FRT Fwd, cctgtaaggagggctccat; Seq FRT Rev, atgagtctgtagggatgct.

Generation of CD11c<sup>ACLEC-2</sup> was achieved by crossing *Clec1b*<sup>fl/fl</sup> with CD11c-Cre mice (B6.Cg-Tg(Igax-cre)1.1Reiz; gift from B. Reizis). Both males and females were used for *in vivo* experiments and were aged 8–12 weeks unless otherwise stated in figures. Cre-negative littermates were used as controls in all experiments.

**qPCR analysis of *Clec1b* messenger RNA.** CD11c<sup>+</sup> cells from day 9 cultures of bone marrow in granulocyte-macrophage colony stimulating factor (GM-CSF) (bone marrow dendritic cells and treated with lipopolysaccharide (LPS) for 6 h or from spleens were enriched by MACS using CD11c beads (Miltenyi). RNA was extracted using an RNeasy mini kit (Qiagen) and cDNA generated using Superscript II reverse transcriptase (Invitrogen). qPCR was carried out using Sybr Green comparing *Clec1b* expression to GAPDH in each sample. Primers for *Clec1b*: Fwd, TTTGAGCACAAGTGCAGCCCC; Rev, AAGCAGTTGGTCCACTCTTG.

**Constructs.** PDPN-CFP was as previously described<sup>8</sup>. RhoA and Rac1 FRET biosensors were provided by M. Matsuda<sup>25</sup>. GFP-MLC was as previously described<sup>12</sup>. PDPN wild type and S167A and S167E mutants were cloned into pcDNA3.1-V5-His by PCR using EcoRI and NotI restriction sites.

**Cell lines.** NIH/3T3 fibroblasts were cultured in DMEM plus glucose (Life Technologies, Invitrogen) with 10% fetal bovine serum (FBS) and penicillin and streptomycin (PS). Cells were incubated and maintained at 37 °C in 5% CO<sub>2</sub>. Mouse lymph node FRC lines were generated by first digesting skin-draining lymph nodes from C57BL/6 mice and then culturing adherent stromal cells as previously reported<sup>8</sup>. On day 4, stromal cells were immortalized by infection with HPV-E6-encoding retrovirus and selected with 2.5  $\mu$ M puromycin as for the generation of carcinoma-associated fibroblast cell lines<sup>26</sup>. Immortalized FRCs were isolated by sequential MACS depletion of CD45<sup>+</sup> and CD31<sup>+</sup> cells using biotin-conjugated antibodies and anti-biotin beads (Miltenyi). FRC cell lines and maintained in DMEM plus glucose (Life Technologies, Invitrogen) with 10% FBS, PS and 1% Insulin-Transferrin-Selenium (Life Technologies, Invitrogen) at 37 °C in 5% CO<sub>2</sub>, and split using cell

dissociation buffer (Life Technologies, Invitrogen). Stable knockdown of PDPN in FRCs was achieved with two different short hairpin (sh)RNA lentiviruses obtained from The RNAi Consortium of the Broad Institute that targeted the following sequences: GCTGCATCTTTCTGGATAATA (PDPN KD1) and GTTCTCCCAA CACATCTGAAA (PDPN KD2). FRC cell lines expressing RhoA or Rac1 biosensors were generated by co-transfection of the biosensor plasmid with PiggyBac transposase using lipofectamine 2000, followed by selection with 5  $\mu$ M blasticidin for 2 weeks. GFP-MLC-expressing FRC cell lines were generated by lentiviral transduction and cell sorting. All cell lines were regularly tested for absence of mycoplasma contamination by the Cell Services Laboratory, Cancer Research UK London Research Institute.

**Overexpression studies.** NIH/3T3 cells were plated at a density of 50,000 cells ml<sup>-1</sup> in a glass-bottomed 24-well plate (MatTek) the day before transfection with plasmids encoding GFP, PDPN-CFP or PDPN-V5-His using Effectene transfection reagent (Qiagen) as per supplier's instructions. After 24 h, cells were treated with chemical inhibitors at the concentrations indicated (see Extended Data Fig. 1b) for 6 h before fixation. CLEC-2-Fc was added at 10  $\mu$ g ml<sup>-1</sup> for the time indicated in figures before fixation. In co-transfection experiments, plasmids encoding PDPN-CFP and ezrin-mCherry or PDPN-V5-His and CD44-GFP were added in equal amounts to the transfection mix. The cells were analysed by fluorescence microscope ( $\times 20$ , Nikon Eclipse Te2000-S) and contraction status scored manually. Cells were grouped into either contracted, partially contracted, spread or collapsed, depending on their morphology. High-resolution images were taken using a confocal microscope (Zeiss 710 using a  $\times 20/0.8$ NA objective).

**siRNA knockdown of RhoGEFs.** siRNA smartpools containing the following sequences were obtained from Dharmacon. NIH/3T3 cells or FRC cell lines were transfected using Dharmafect reagent 2 according to the manufacturer's instructions. Mef2 (Dbl), UGAUCAGUCUCCCAAAUUG; CCAUGCCUUUCAUCAA UUA; GGUGUAUACCGCAAAUUG; CAAAGUGCAUAGACUCUUA; GEF-H1 (ARHGEF2), CAACAUUGCUGGACAUUUG; GCACUGGAUGCUGGA AGA; GUACCAAGGUCAAGCAGAA; UGGAAUCCCUAAUUGAUGA; LARG (ARHGEF12), GAUCAAGUCUGCCAGAAA; GGACGGAGCUGUAUUUG CA; GAAAGGAGUUCACAAUUG; GAAAGGAGUUCACAAUUG; p115 rhogef (ARHGEF1), GGUGUAACCUCACUCUGA; GGAAAGACCGAGGC AACUA; GGCAAGAGGUCAUCAGCGA; GGGCUGAGCAGUAUCCUAG.

**Immunofluorescence staining.** Cells were fixed with 4% paraformaldehyde (PFA) in PBS for 10 min before permeabilization in PBS containing 0.2% Triton-X for 10 min at room temperature. The cells were stained with 4',6-diamidino-2-phenylindole (DAPI; Sigma, d9542) to reveal DNA in cell nuclei and tetramethylrhodamine (TRITC)-phalloidin (Sigma, p1951) to reveal F-actin in 3% bovine serum albumin with PBS plus 0.1% Tween-20. Cells were stained using anti-pMLC (S19) (Cell Signaling Technology #3675) or anti-pERM antibody (Cell Signaling Technology #3141) followed by appropriate Alexafluor-conjugated secondary antibodies (Invitrogen). PDPN-V5 was stained using anti-V5 fluorescein isothiocyanate (FITC)-conjugated antibody (Invitrogen R963-25). ARP2/3 was stained using anti-p34-Arc/ARPC2 antibody (07-227 Millipore).

**Lipid raft co-localization analysis.** Lipid raft labelling was carried out on ice on unfixed cells according to manufacturer's instructions (Vybrant lipid raft labelling kit 555, Invitrogen). Cells were then fixed in 4% paraformaldehyde before staining for PDPN (8.1.1 AF660-conjugated; eBioscience 50-5381-80). Confocal immunofluorescence images ( $\times 63/1.4$ NA oil immersion objective) of FRCs plated on glass either untreated or treated with 10  $\mu$ g ml<sup>-1</sup> CLEC-2-Fc for 30 min before staining were analysed using Zen image analysis software (Zeiss). The correlation coefficient *R* was calculated based on pixel intensity.

**CLEC-2-Fc treatment *in vitro*.** Generation of CLEC-2-Fc was as previously described<sup>8</sup>. Soluble CLEC-2 was used at 10  $\mu$ g ml<sup>-1</sup> diluted in cell culture medium. For CLEC-2-beads, 10  $\mu$ m protein A-coated microspheres (Bangs laboratories) were incubated with 100  $\mu$ g ml<sup>-1</sup> CLEC-2-Fc diluted in PBS for 1 h at 4 °C then washed with cell culture medium. CLEC-2-Fc or CLEC-2-beads were added for the time indicated in the figures before fixation.

**Lymph node expansion *in vivo*.** Mice were immunized with 100  $\mu$ l of an emulsion of OVA in CFA (100  $\mu$ g OVA per mouse) or PBS in incomplete Freund's adjuvant (IFA) subcutaneously in the right flank. Draining inguinal lymph nodes were taken for analysis at the indicated days after immunization; left-side non-draining lymph nodes were taken for comparison. Where mice were treated with CLEC-2-Fc, they received 10  $\mu$ l (1  $\mu$ g in PBS) subcutaneously adjacent to site of immunization on days 1 and 3. Lymph nodes were first weighed, then digested as previously described<sup>27</sup>. Cells were counted and stained for flow cytometry analysis. Alternatively, intact lymph nodes were fixed in 10% formalin for immunohistochemistry analysis.

**Lymph node deformation assay.** Draining and non-draining inguinal lymph nodes were taken from CD11c<sup>ACLEC-2</sup> mice and littermate controls on day 5 after CFA immunization. Lymph nodes were placed on the contact points of a digimatic



thickness gauge (Mitutoyo) and subjected to 1.4 N of applied force. Deformability was calculated as follows: Deformability =  $1 - (\text{lymph node size under } 1.4 \text{ N} / \text{lymph node size before force was applied})$ . Each lymph node was measured twice and the results averaged.

**Flow cytometry.** Cells isolated from lymph nodes, FRCs or NIH/3T3 cell lines were suspended in FACS buffer (PBS 2% FCS, 2 mM EDTA) and first blocked with anti-CD16/CD32 (eBioscience) for all staining procedures. Cells were counted on a FACS Calibur by reference to fluorescence beads. Stained cells were analysed on either a FACS Calibur or LSRII. Antibodies used for staining: CD45.1 (BD Biosciences), CD140a (clone APA5 eBioscience), CD31 (BD Biosciences), PDPN (clone 8.1.1 eBioscience), CD35 (clone 8C12 BD Biosciences), VCAM-1 (clone M/K-2 Abcam), CD44 (clone IM7 BD Biosciences), CD3 (BD Biosciences), CD19 (clone ID3 BD Biosciences), MHCII (I-A/I-E BD Biosciences), CD11c (clone HL3 BD Biosciences).

**Immunohistochemistry.** Tissue sections 4  $\mu\text{m}$  thick were cut at three levels (with 150  $\mu\text{m}$  between the levels) from formalin-fixed paraffin-embedded lymph nodes. These sections were then stained using the Vector ABC elite detection system (PK6100 1:250, Vector Laboratories). Slides were first incubated with a primary antibody for 1 h (anti-GFP: AB6673 1:350, Abcam; anti-PDPN: DM3501, Acris), followed by incubation with biotinylated-labelled secondary antibody (1:250, Vector Laboratories) for 45 min. ABC complex was then applied for 30 min and staining was completed by 3 min incubation with DAB and chromogen (SK 4100, Vector Laboratories). Slides were counterstained with haematoxylin and mounted. ER-TR7 staining was conducted on 10- $\mu\text{m}$ -thick frozen lymph node sections (Santa Cruz, sc-73355-AF488).

**Gap analysis.** Quantification of gap sizes was carried out in MATLAB. PDPN signal was isolated and converted to greyscale. The images were thresholded, background subtracted, small objects removed, then converted to binary images using ImageJ software. A circle-fitting algorithm was applied whereby, in each step, the largest circle that could fit in the gaps and that did not overlap with other fitted circles was recorded. The distribution of radii of circles that fill the image was then weighted according to area such that larger circles had a proportionally greater weighting than smaller circles. The plot of distribution of radii was smoothed to transform the data from a discrete pixel size distribution to a continuous micrometre-size distribution. Raw data were analysed using Fisher's exact test ( $P = 4.194 \times 10^{-9}$ ) to determine differences in circles with radius  $>15 \mu\text{m}$ . Each analysis was performed on images of the FRC network from  $>8$  individual mice per group. No images were excluded from the analysis and all data were combined and are represented in the graphs shown (16 images per group). MATLAB script is available upon request.

**Automated morphometric analysis.** The slides were digitized with a commercial image analysis system (Ariol; Leica Biosystems). The program was trained to recognize GFP<sup>+</sup> stained nuclei by size, shape and staining intensity. T-cell areas were identified by density of haematoxylin staining and traced out manually on

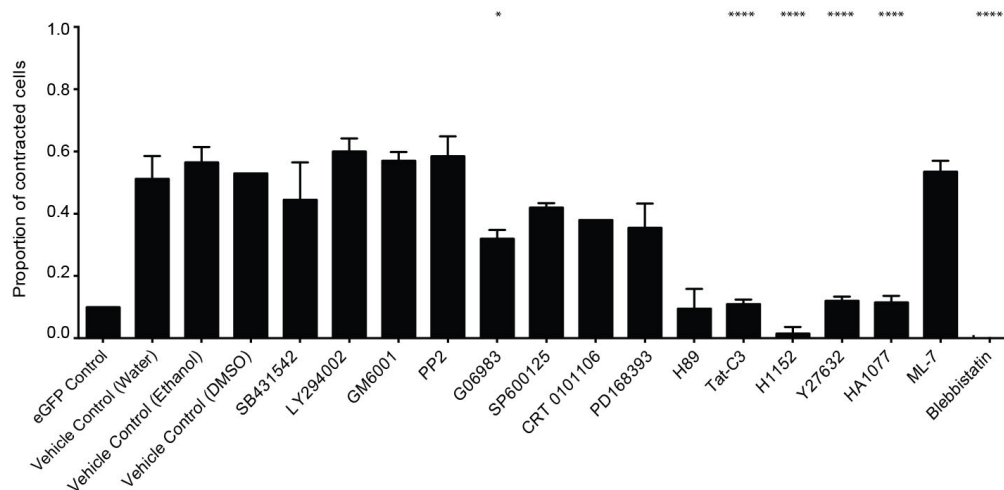
each lymph node in the scan. Automated analysis then counted the total number of GFP<sup>+</sup> nuclei in each T-cell area (24 areas analysed per group). Area/FRC nuclei were compared using Mann–Whitney *U*-test.

**Rac1-GTP pulldown assay.** FRC cell lines either treated with CLEC-2-Fc for 30 min or stably depleted of PDPN (shRNA) were subjected to Rac1 pulldown and analysis as per manufacturer's instructions (kit 16118 Pierce). Rac1 levels in pulldowns and whole lysate were compared by western blot.

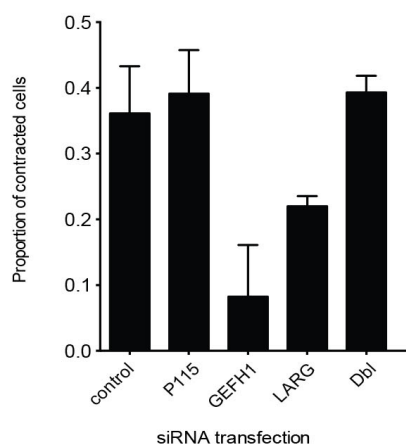
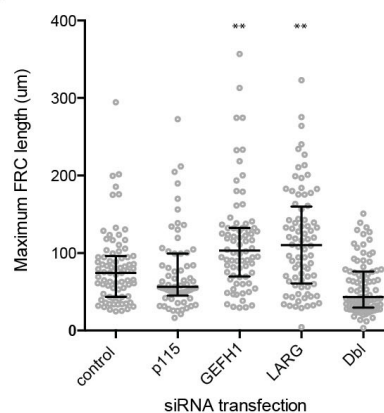
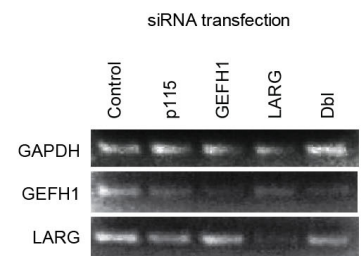
**Three-dimensional cell culture and gel contraction assay.** Control or PDPN knock-down FRCs were seeded at 10,000 per well in 150  $\mu\text{l}$  collagen/matrigel matrix<sup>8,26,28</sup>. Gels were set at 37 °C for 30 min then covered with cell culture medium. In some wells, the following were added to both the gel mix and medium: 10  $\mu\text{g ml}^{-1}$  CLEC-2-Fc, 10  $\mu\text{M}$  ROCK inhibitor (Y27632) or 10  $\mu\text{g ml}^{-1}$  anti-PDPN antibody (R&D, AF3244). Contraction of the gel at day 3 was quantified as the ratio of contracted gel/original area and plotted relative to control. Gels were stained with TRITC-labelled phalloidin and DAPI for maximum length analysis (Imaris). The furthest points of each individual cell were measured in *x,y,z* coordinates and vectors calculated for comparison (Extended Data Fig. 5).

**Statistics.** Sample size for *in vivo* experiments was determined by litter size as littermate controls were used in all comparisons. For *in vitro* experiments, all cells within the sample were scored and none excluded and experiments were repeated at least once to ensure reproducibility and adequate statistical power. Category data from overexpression studies were analysed using Fisher's exact test (R software). Changes to contraction and morphology were analysed using Mann–Whitney *U*-tests. Analysis of cell populations during a time course of immunization was conducted with one-way ANOVA and followed by Kruskal–Wallis multiple comparisons test. Comparison of lymph node size and cellularity between control and CD11c<sup>ΔCLEC-2</sup> across different treatment groups was analysed using two-way ANOVA followed by multiple comparisons test between treatments and genotypes. Appropriate statistical tests were chosen for the data set following advice from a mathematician (R.P.J.). Chemical screen for cell contraction inhibitors was scored by an independent observer (A.J.F.) who was unbiased as to the predicted results.

25. Hirata, E. *et al.* *In vivo* fluorescence resonance energy transfer imaging reveals differential activation of Rho-family GTPases in glioblastoma cell invasion *J. Cell Sci.* **125**, 858–868 (2012).
26. Calvo, F. *et al.* Mechanotransduction and YAP-dependent matrix remodelling is required for the generation and maintenance of cancer-associated fibroblasts. *Nature Cell Biol.* **15**, 637–646 (2013).
27. Fletcher, A. L. *et al.* Reproducible isolation of lymph node stromal cells reveals site-dependent differences in fibroblastic reticular cells. *Front. Immunol.* **2**, 35 (2011).
28. Gaggioli, C. *et al.* Fibroblast-led collective invasion of carcinoma cells with differing roles for RhoGTPases in leading and following cells. *Nature Cell Biol.* **9**, 1392–1400 (2007).

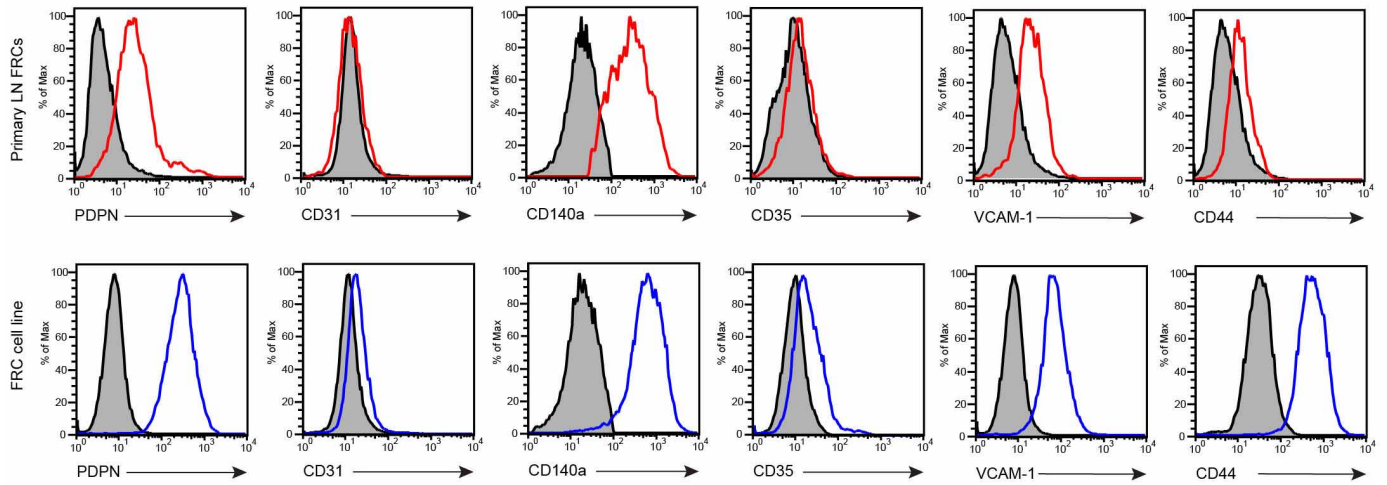
**a****b**

Inhibitor	Target	Buffer	Final Conc.
SB431542	Transforming growth factor-beta receptor	DMSO	10 $\mu$ M
LY294002	Phosphoinositide 3-kinase (PI3K)	ETOH	20 $\mu$ M
GM6001	Matrix metalloproteinase (MMP)	H <sub>2</sub> O	10 $\mu$ M
PP2	Src	H <sub>2</sub> O	10 $\mu$ M
G06983	Protein kinase C (PKC)	DMSO	10 $\mu$ M
SP600125	Jun N-terminal kinase	DMSO	10 $\mu$ M
CRT 0101106	LIM Kinase	DMSO	10 $\mu$ M
H89	Protein Kinase A (PKA)/Rho kinase (ROCK)	H <sub>2</sub> O	10 $\mu$ M
Tat-C3	RhoA	H <sub>2</sub> O	10 $\mu$ M
H1152	Rho Kinase (ROCK)	H <sub>2</sub> O	10 $\mu$ M
Y27632	Rho Kinase (ROCK)	H <sub>2</sub> O	10 $\mu$ M
HA1077	Rho Kinase (ROCK)	DMSO	10 $\mu$ M
ML-7	MLC kinase	H <sub>2</sub> O	10 $\mu$ M
Blebbistatin	Myosin II	H <sub>2</sub> O	10 $\mu$ M

**c****d****e**

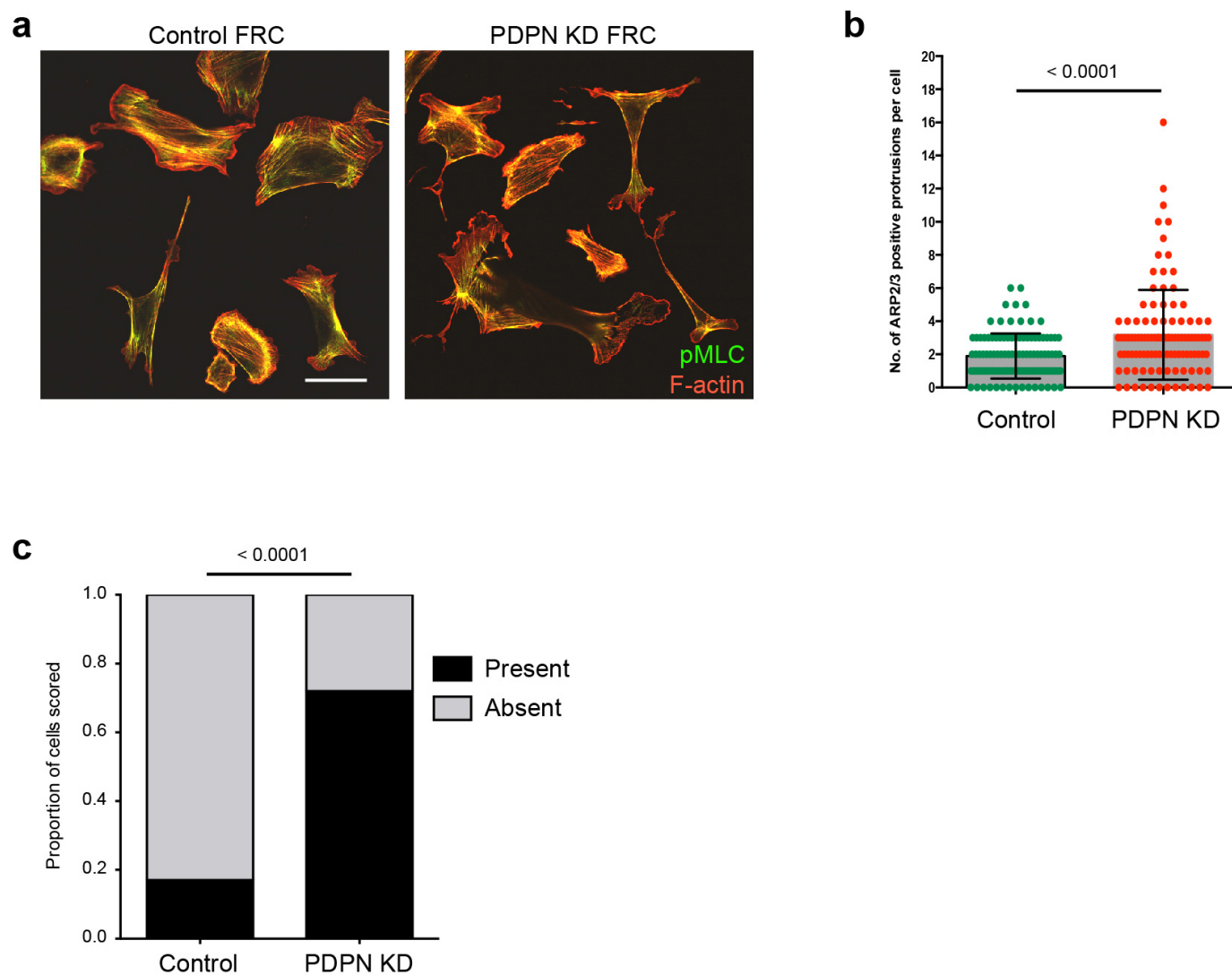
**Extended Data Figure 1 | Screen for inhibitors of PDPN-mediated cell contractility.** **a**, Quantification of proportion of contracted NIH/3T3 fibroblasts expressing enhanced (e)GFP control or PDPN-CFP and treated with the indicated inhibitors or vehicles. Statistically significant inhibition: \*\*\*\* $P < 0.00001$  and \* $P = 0.01$ , Fisher's exact test. Data represent mean  $\pm$  s.d. of three independent experiments. **b**, Chemical inhibitors used in **a** and their targets. **c**, Contraction score of PDPN-expressing NIH/3T3 fibroblasts transfected with siRNA smartpools targeting the indicated Rho GEFs

(MU-046870-01-0002, MU-040120-00-0002, MU-047092-01-0002, MU-041056-01-0002, Dharmacon, GE Healthcare). \*\* $P < 0.05$ , one-way ANOVA. **d**, Maximum length of FRCs in collagen gels measured in three dimensions from 100- $\mu$ m-deep confocal z-stacks. Each point represents one FRC. \*\* $P < 0.05$ , one-way ANOVA. **e**, PCR analysis of *Rho GEF* mRNA expression in FRC cell lines after siRNA knockdown in comparison to expression of GAPDH.



**Extended Data Figure 2 | Generation of FRC lines.** Comparison of an FRC cell line generated by immortalization of primary FRCs (bottom, blue) with primary FRCs in lymph node (LN) cell suspensions cultured for 7 days (top, red). Grey histograms indicate isotype-matched control antibody staining.

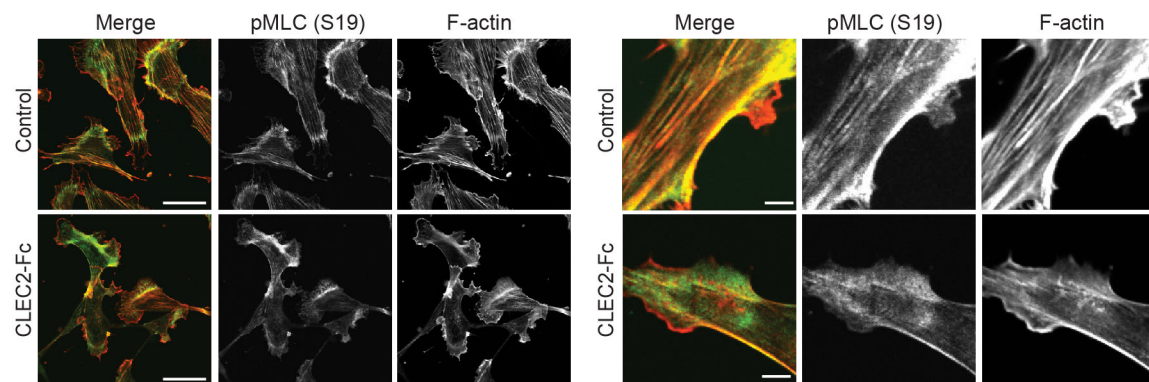
Histograms of primary lymph node cultures are gated on CD45<sup>-</sup> CD140a<sup>+</sup> cells to exclude haematopoietic cells and other stromal subsets. Histograms of the FRC cell line are gated only on live cells.



**Extended Data Figure 3 | Loss of PDPN results in FRC spreading and actin polymerization.** **a**, Single optical slice (1  $\mu\text{m}$ ) showing morphology and pMLC organization of control and PDPN knockdown (KD) FRCs. pMLC (S19) (green), F-actin (red). Scale bar, 50  $\mu\text{m}$ . **b**, Quantification of the number of ARP2/3-positive protrusions per FRC comparing control (green) and PDPN

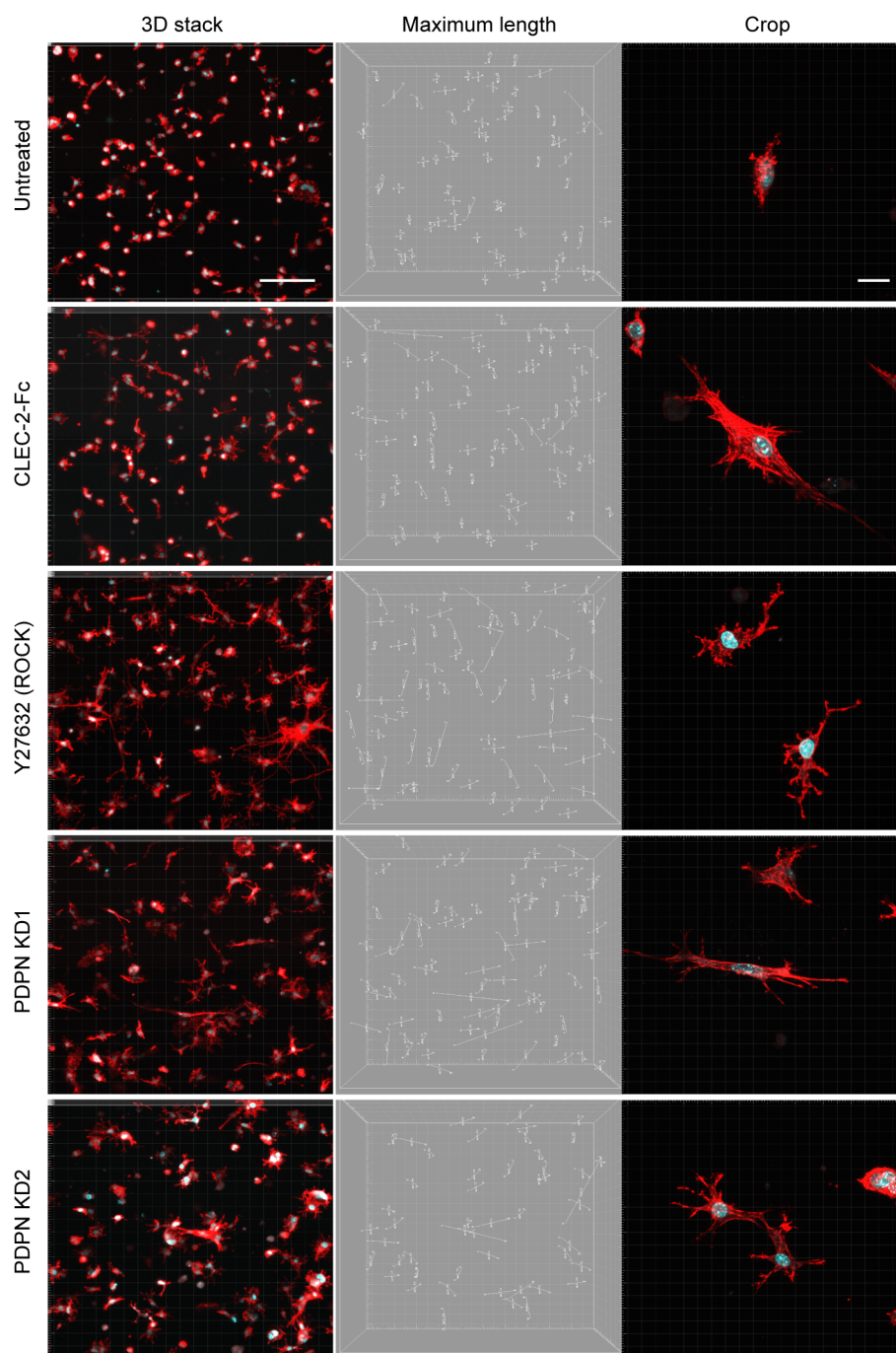
knockdown (red) cell lines. Data represent mean  $\pm$  s.d., each point represents an individual FRC.  $P < 0.0001$ , Mann-Whitney  $U$ -test. **c**, Quantification of tail retraction defects comparing control and PDPN knockdown FRCs. Data are collated from  $>100$  cells;  $P < 0.0001$ , Fisher's exact test. 'Present' means that tail retraction defects were deemed present by the observer.





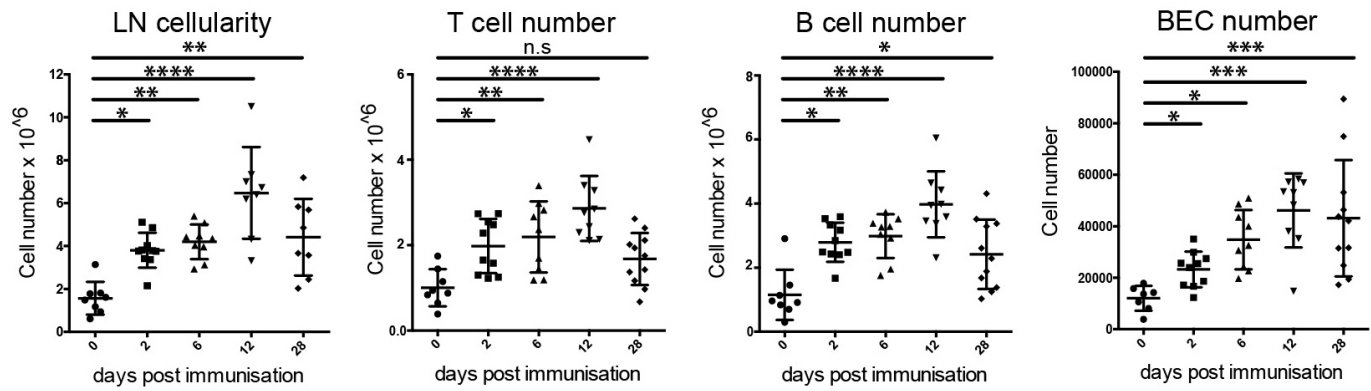
**Extended Data Figure 4 | Loss of pMLC and F-actin filaments after treatment of FRCs with CLEC-2.** Single optical slices ( $1\ \mu\text{m}$ ) of FRC cell lines treated for 30 min with  $10\ \mu\text{g ml}^{-1}$  soluble CLEC-2-Fc protein, fixed and

stained for pMLC (S19) (green) and F-actin (red). Scale bars,  $50\ \mu\text{m}$ . Higher magnification shown to the right. Scale bars,  $5\ \mu\text{m}$ .



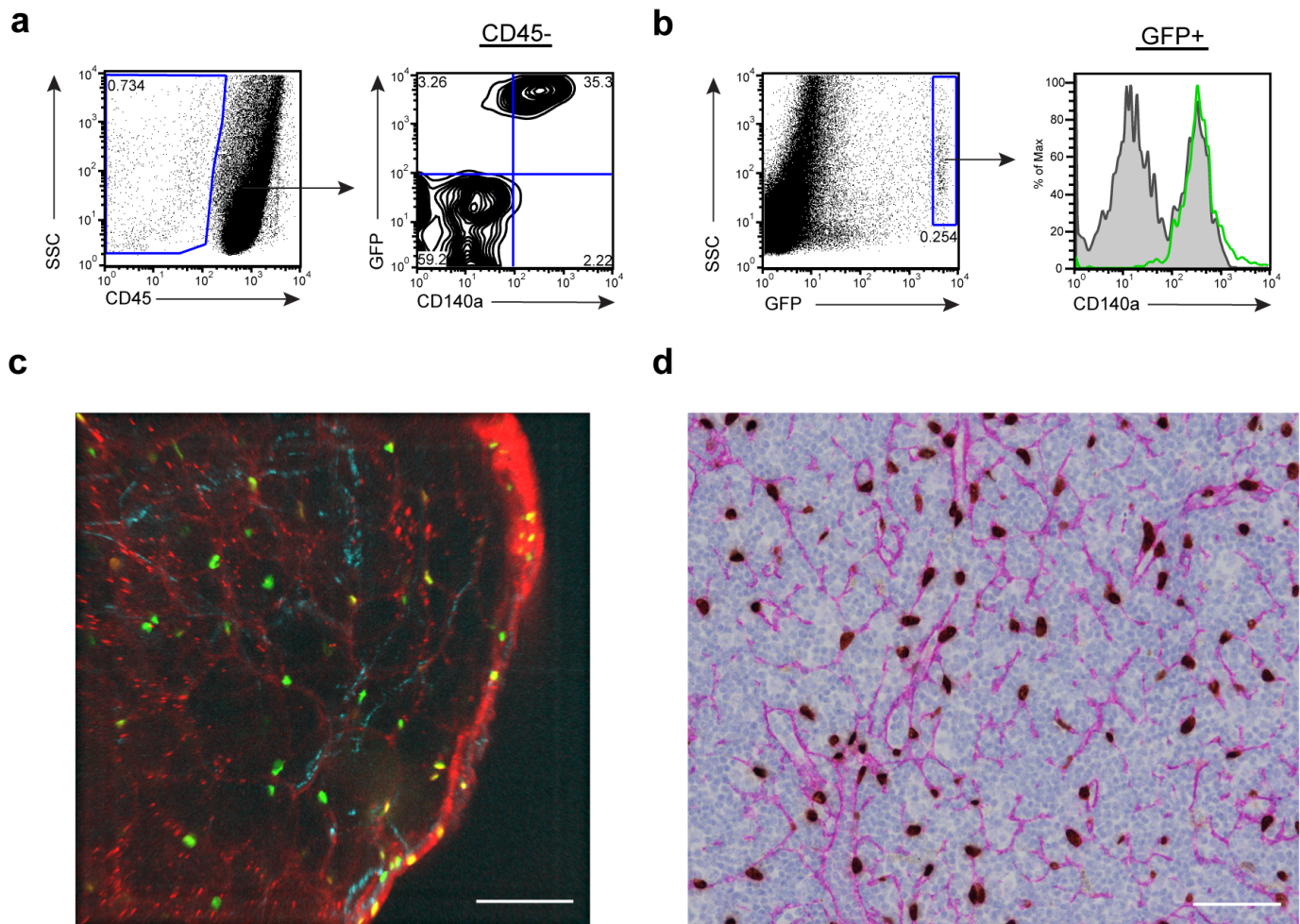
**Extended Data Figure 5 | Elongated morphology of CLEC-2-treated FRCs in three-dimensional cultures.** Quantification of maximum cell length for 100- $\mu$ m-deep z-stacks. FRCs cultured in three-dimensional collagen/matrigel matrix for 3 days treated with CLEC-2-Fc, ROCK inhibitor 10  $\mu$ M (Y27632), or stably knocked down (KD) for PDPN expression. Left, projection of the

three-dimensional stack; staining F-actin (red), DNA (cell nuclei, blue). Scale bar, 200  $\mu$ m. Centre,  $x,y,z$  coordinates and length of each vector for each end of each cell in three dimensions as quantified using Imaris image analysis software. Right, example of cell morphology in each treatment group; staining F-actin (red), DNA (cell nuclei, blue). Scale bar, 50  $\mu$ m.



**Extended Data Figure 6 | Time course of lymph node expansion after OVA/CFA immunization.** Total cellularity, number of T cells ( $CD45^+$ ,  $CD3^+$ ), B cells ( $CD45^+$ ,  $CD19^+$ ) and biliary epithelial cells (BECs;  $CD45^-$   $PDPN^-$   $CD31^+$ ) in draining lymph nodes (LNs) at different times after OVA/CFA immunization. Each point represents one lymph node and data

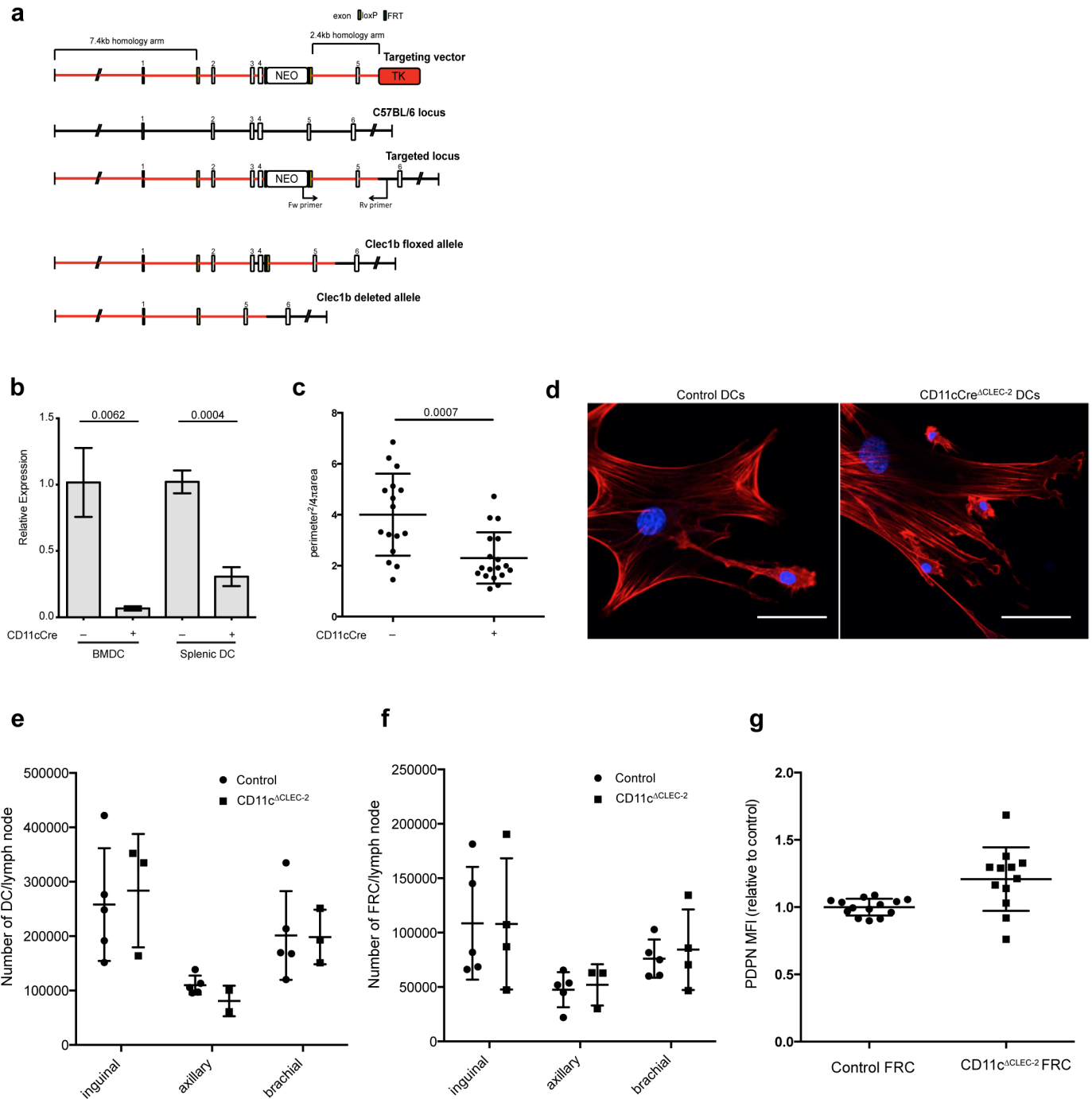
show mean  $\pm$  s.d. of two independent experiments scoring 8–12 mice. Day 0 represents lymph nodes from non-immunized mice. \* $P < 0.05$ , \*\* $P > 0.001$ , \*\*\* $P > 0.0001$ , \*\*\*\* $P < 0.00001$ , differences between non-immunized and immunized mice as calculated using one-way ANOVA test followed by Tukey's multiple comparisons test.



**Extended Data Figure 7 | FRCs are selectively labelled in PDGFR $\alpha$ KI-H2B-GFP mice.** **a**, Analysis of skin draining lymph nodes from PDGFR $\alpha$ KI-H2B-GFP mice showing lymph node stromal cells co-expressing CD140a (PDGFR $\alpha$ ) and GFP. Left, gate for CD45<sup>-</sup> stroma; right, GFP and CD140a expression of CD45<sup>-</sup> gate. **b**, Flow cytometry analysis showing that GFP<sup>+</sup> cells are CD140a<sup>+</sup>. Left, gating for GFP<sup>+</sup> lymph node cells; right, CD140a expression of GFP<sup>+</sup> gate (green) compared with CD45<sup>-</sup> cells (grey). **c**, z-Stack

of PDGFR $\alpha$ KI-H2B-GFP lymph node imaged *ex vivo* using two-photon microscopy. FRC nuclei (green), second harmonic signal (collagen) (blue). Wheat germ agglutinin AF647 (red) was injected subcutaneously 5 min before lymph node extraction to label conduits. Scale bar, 200  $\mu$ m. **d**, Immunohistochemical staining of paraffin-embedded sections of lymph nodes from PDGFR $\alpha$ KI-H2B-GFP mice. Staining GFP (brown) and PDPN (pink), counterstained with haematoxylin (blue). Scale bar, 200  $\mu$ m.





**Extended Data Figure 8 | Generation and characterization of  $CD11c^{\Delta CLEC-2}$  mice.** **a**, Scheme of targeting approach to allow conditional deletion of *Clec1b* exons 2, 3 and 4. *loxP* sites are shown in yellow. **b**, *Clec1b* mRNA in lipopolysaccharide (LPS)-treated bone marrow dendritic cells (BMDCs) or freshly isolated  $CD11c^{+}$  splenocytes from  $CD11c^{\Delta CLEC-2}$  mice and  $Cre^{neg}$  littermates. Data are represented as relative expression compared to control and depict mean  $\pm$  s.d. from six replicates from two independent experiments. *P* values were calculated using Student's *t*-test. **c**, Quantification of bone marrow-derived dendritic cell (DC) morphology cultured in contact with FRCs. Data indicate score of  $perimeter^2/(4\pi \times area)$ , with area and perimeter

calculated from immunofluorescence imaging using ImageJ software. Higher scores indicate increased elongation and/or protrusions. *P* = 0.0007, Mann-Whitney *U*-test. **d**, Representative images from **c** showing dendritic cells spreading over FRCs. F-Actin (red), cell nuclei (blue). Scale bar, 20  $\mu$ m. **e**, **f**, Total dendritic cell numbers (**e**) and total FRC numbers (**f**) in steady-state skin draining lymph nodes of control versus  $CD11c^{\Delta CLEC-2}$  mice. Each data point represents one lymph node. **g**, PDPN surface expression by FRCs from control and  $CD11c^{\Delta CLEC-2}$  mice as measured by flow cytometry and represented relative to the control group. MFI, mean fluorescence intensity.

# Diabetes recovery by age-dependent conversion of pancreatic $\delta$ -cells into insulin producers

Simona Chera<sup>1</sup>, Delphine Baronnier<sup>1</sup>, Luiza Ghila<sup>1</sup>, Valentina Cigliola<sup>1</sup>, Jan N. Jensen<sup>2</sup>, Guoqiang Gu<sup>3</sup>, Kenichiro Furuyama<sup>1</sup>, Fabrizio Thorel<sup>1</sup>, Fiona M. Gribble<sup>4</sup>, Frank Reimann<sup>4</sup> & Pedro L. Herrera<sup>1</sup>

**Total or near-total loss of insulin-producing  $\beta$ -cells occurs in type 1 diabetes<sup>1,2</sup>. Restoration of insulin production in type 1 diabetes is thus a major medical challenge. We previously observed in mice in which  $\beta$ -cells are completely ablated that the pancreas reconstitutes new insulin-producing cells in the absence of autoimmunity<sup>3</sup>. The process involves the contribution of islet non- $\beta$ -cells; specifically, glucagon-producing  $\alpha$ -cells begin producing insulin by a process of reprogramming (transdifferentiation) without proliferation<sup>3</sup>. Here we show the influence of age on  $\beta$ -cell reconstitution from heterologous islet cells after near-total  $\beta$ -cell loss in mice. We found that senescence does not alter  $\alpha$ -cell plasticity:  $\alpha$ -cells can reprogram to produce insulin from puberty through to adulthood, and also in aged individuals, even a long time after  $\beta$ -cell loss. In contrast, before puberty there is no detectable  $\alpha$ -cell conversion, although  $\beta$ -cell reconstitution after injury is more efficient, always leading to diabetes recovery. This process occurs through a newly discovered mechanism: the spontaneous *en masse* reprogramming of somatostatin-producing  $\delta$ -cells. The juveniles display 'somatostatin-to-insulin'  $\delta$ -cell conversion, involving dedifferentiation, proliferation and re-expression of islet developmental regulators. This juvenile adaptability relies, at least in part, upon the combined action of FoxO1 and downstream effectors. Restoration of insulin producing-cells from non- $\beta$ -cell origins is thus enabled throughout life via  $\delta$ - or  $\alpha$ -cell spontaneous reprogramming. A landscape with multiple intra-islet cell interconversion events is emerging, offering new perspectives for therapy.**

To determine how ageing affects the mode and efficiency of  $\beta$ -cell reconstitution after  $\beta$ -cell loss, we administered diphtheria toxin (DT) to adult (2-month-old) or aged (1- and 1.5-year-old) *RIP-DTR* mice, whose  $\beta$ -cells bear DT receptors<sup>3</sup>, and followed them for up to 14 months. Collectively, we found that  $\alpha$ -to- $\beta$ -cell conversion is the main mechanism of insulin cell generation after massive  $\beta$ -cell loss in adult post-pubertal mice, whether middle-aged or very old, and that  $\alpha$ -cells are progressively recruited into insulin production with time (Extended Data Fig. 1 and Supplementary Tables 1–5).

We focused on regeneration potential during early postnatal life by inducing  $\beta$ -cell ablation before weaning, at 2 weeks of age (Fig. 1a). We found that prepubescent mice rapidly recover from diabetes after near-total  $\beta$ -cell loss: 4 months later all mice were almost normoglycaemic, thus displaying a faster recovery relative to adults (Fig. 1b and Extended Data Fig. 2a, b; see also Extended Data Fig. 1a).

Histologically, 99% of  $\beta$ -cells were lost at 2 weeks after DT administration (Fig. 1c). The  $\beta$ -cell number increased by 45-fold 4 months after ablation, representing 23% of the normal age-matched  $\beta$ -cell mass (Fig. 1c and Supplementary Table 6) and correlating with recovery of normoglycaemia<sup>1</sup>.

All animals remained normoglycaemic for the rest of their life (Supplementary Table 6). Mice were neither intolerant to glucose nor insulin resistant during the period of analysis, up to 15 months after injury (Extended Data Fig. 2c–e).

We investigated whether the new insulin<sup>+</sup> cells were reprogrammed  $\alpha$ -cells, as in adults, using *glucagon-rtTA*; *TetO-Cre*; *R26-YFP*; *RIP-DTR* pups (Fig. 1d). We observed that almost no insulin<sup>+</sup> cells co-expressed yellow fluorescent protein (YFP) or glucagon (Supplementary Table 7), indicating that  $\alpha$ -cells do not reprogram in juveniles.

We further explored the age-dependency of rescue after near-total  $\beta$ -cell loss. To this aim, normoglycaemic 5-month-old mice, which had recovered from  $\beta$ -cell loss at 2 weeks of age, were re-administered DT to ablate the regenerated insulin<sup>+</sup> cells. One month following the second ablation, 30% of the insulin-containing cells also contained glucagon (Extended Data Fig. 2f and Supplementary Table 8), like  $\beta$ -cell-ablated adults (Extended Data Fig. 1k), confirming that the pre-pubertal regeneration mechanism is restricted temporally.

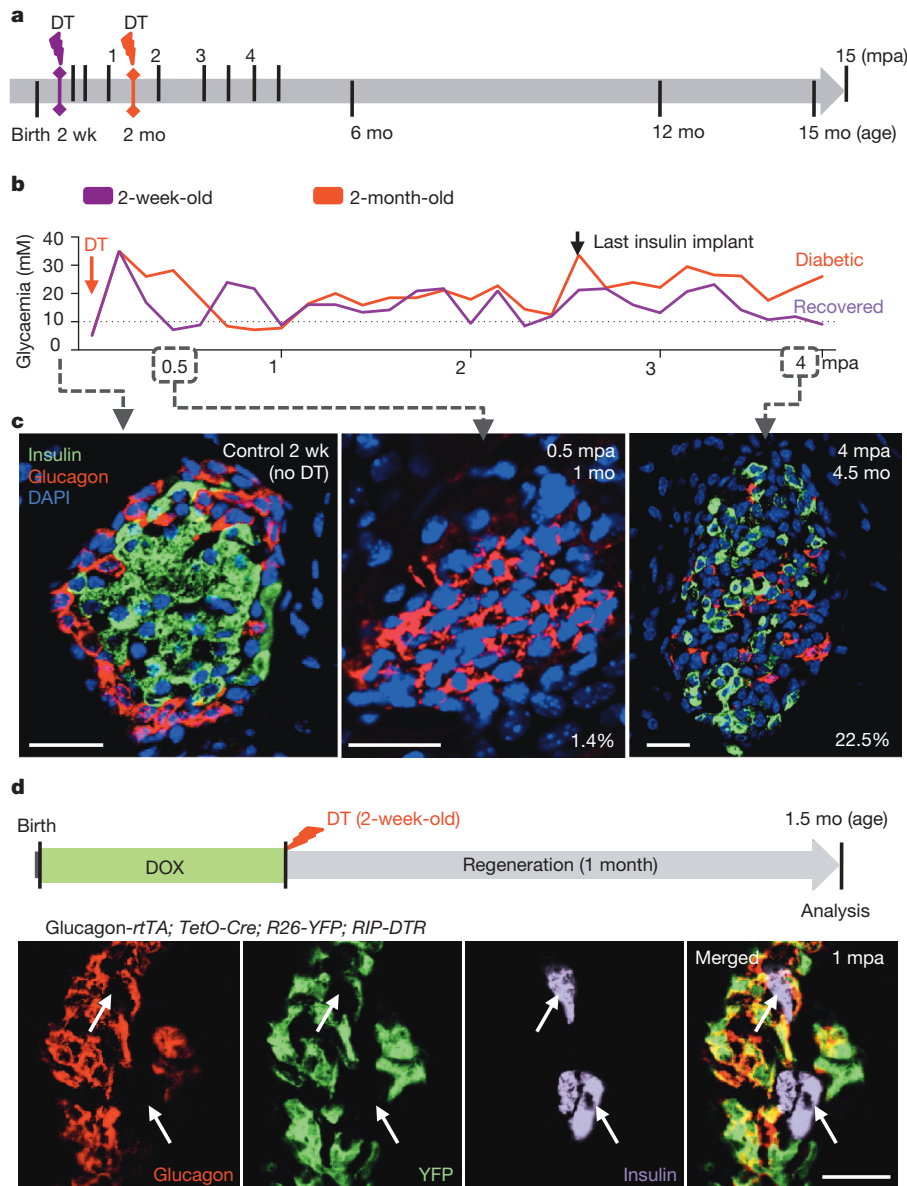
We measured proliferation rates at different time-points over 2 months of regeneration. The proportion of Ki67-labelled insulin<sup>+</sup> cells was very low (Extended Data Fig. 2g and Supplementary Table 9), indicating that neither escaping  $\beta$ -cells nor regenerated insulin<sup>+</sup> cells proliferate during this period. However, there was a transient 3.5-fold increase in the number of insular Ki67<sup>+</sup> cells 2 weeks after ablation, unlike in adult animals (Extended Data Fig. 2h and Supplementary Table 10). Replicating cells were hormone-negative, chromogranin-A-negative, and were not lineage traced to either  $\alpha$ - or escaping  $\beta$ -cells (Extended Data Fig. 2i, j).

Coincident with the peak of islet cell proliferation, we noticed in pups a 4.5-fold decrease in the number of somatostatin (Sst)-producing  $\delta$ -cells (from 13 to 3  $\delta$ -cells per islet section; Extended Data Fig. 3a and Supplementary Table 11) and a 76-fold decrease of *Sst* transcripts (Extended Data Fig. 3b), without any indication of increased islet cell death. We therefore lineage traced  $\delta$ -cells and observed that regenerated insulin-producing cells were dedifferentiated  $\delta$ -cells. At 2 months of age in *Sst-Cre*; *R26-YFP*; *RIP-DTR* mice, about 81% of  $\delta$ -cells were YFP<sup>+</sup> in the absence of  $\beta$ -cell ablation, whereas  $\alpha$ - and  $\beta$ -cells were labelled at background levels (0.9% for  $\beta$ -cells and 0.2% for  $\alpha$ -cells; Extended Data Fig. 3c, d and Supplementary Table 12). During  $\beta$ -cell reconstitution in pups, 2 weeks after  $\beta$ -cell ablation, 80% of YFP<sup>+</sup> cells were proliferating (Ki67<sup>+</sup>) and Sst-negative (Fig. 2a, b and Supplementary Table 13), while most Ki67<sup>+</sup> cells were YFP-labelled (85%; Supplementary Table 14).

These observations suggest that in  $\beta$ -cell-ablated pre-pubertal mice most  $\delta$ -cells undergo a loss of Sst expression and enter the cell cycle.

We further investigated the fate of proliferating dedifferentiated  $\delta$ -cells. At 1.5 months post-ablation, most insulin<sup>+</sup> cells expressed YFP (90%), indicating their  $\delta$ -cell origin (Fig. 2c, d and Supplementary Table 15). Furthermore, in contrast to non-ablated age-matched controls, where all YFP<sup>+</sup> cells were Sst<sup>+</sup> (>99%), about half of YFP<sup>+</sup> cells were insulin<sup>+</sup> after 1.5 months of regeneration (45%; Fig. 2e and Supplementary Table 16). This reveals that half of the progeny of dedifferentiated  $\delta$ -cells becomes insulin expressers. Bihormonal Sst<sup>+</sup>/insulin<sup>+</sup> cells were rare (Supplementary Table 17).

<sup>1</sup>Department of Genetic Medicine & Development, Faculty of Medicine, University of Geneva, 1 rue Michel-Servet, 1211 Geneva-4, Switzerland. <sup>2</sup>Novo Nordisk A/S, Niels Steensens Vej 6, DK-2820 Gentofte, Denmark. <sup>3</sup>Cell and Developmental Biology, Vanderbilt University Medical Center, 465 21st Av. South, Nashville, Tennessee 37232, USA. <sup>4</sup>Cambridge Institute for Medical Research, Hills Road, Cambridge CB2 0XY, UK.



**Figure 1 |  $\beta$ -cell ablation before puberty and diabetes recovery.**

**a**, Experimental designs depicting the ages at DT administration and the various analyses. mpa, months post-ablation. **b**, Comparative evolution of glycaemia in  $\beta$ -cell-ablated juveniles ( $n = 5$ ) and middle-aged adults ( $n = 4$ ); 2.5 months after  $\beta$ -cell ablation, insulin administration was stopped ( $P = 0.0014$ , Mann–Whitney test). Dotted line shows upper limit of

normoglycaemia (10 mM). **c**, Islets from 2-week-old mice with no DT treatment (control), 1-month-old mice at 0.5 mpa and 4.5-month-old mice at 4 mpa (Supplementary Table 6). Percentages refer to  $\beta$ -cell mass relative to age-matched unablated controls. DAPI, 4',6-diamidino-2-phenylindole. **d**,  $\alpha$ -Cell tracing in pups. DOX, doxycycline. Scale bars, 20  $\mu$ m.

Combined, these observations show that at the cell population level, each dedifferentiated  $\delta$ -cell yields one insulin-expressing cell and one  $Sst^+$  cell (Extended Data Fig. 4).

We confirmed with two other assays that regeneration and diabetes recovery in juvenile mice are  $\delta$ -cell-dependent: by inducing  $\beta$ -cell destruction with streptozotocin (STZ) instead of DT (Extended Data Fig. 5a–c), and by co-ablating  $\beta$ - and  $\delta$ -cells simultaneously in *Sst-Cre; R26-YFP; RIP-DTR* pups. In the absence of  $\delta$ -cells there was no insulin $^+$  cell regeneration, and no recovery (Fig. 2f).

In adults,  $\delta$ -cells neither dedifferentiated nor proliferated after  $\beta$ -cell ablation (Extended Data Fig. 5d, e and Supplementary Table 20). Nevertheless, like  $\alpha$ -cells, a few  $\delta$ -cells reprogrammed into insulin production, so that after 1.5 months of regeneration 17% of the rare insulin-producing cells were YFP $^+$ , that is,  $\delta$ -cell-derived (Extended Data Fig. 5f–h and Supplementary Tables 21, 22).

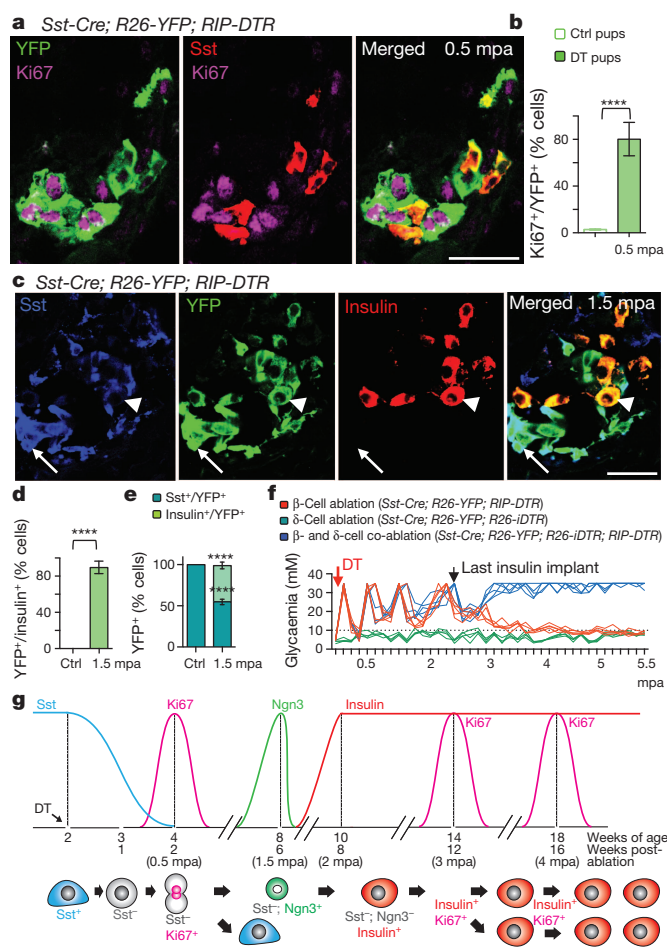
By transplanting *Sst-Cre; R26-YFP; RIP-DTR* juvenile islets into adult wild-type mice we observed that, following  $\beta$ -cell ablation, the newly

formed insulin $^+$  cells were reprogrammed  $\delta$ -cells, thus showing that the pup-specific regeneration is intrinsic to islets (Extended Data Fig. 6).

Contrary to  $\beta$ -cells in age-matched adult mice,  $\delta$ -cell-derived insulin $^+$  cells replicated transiently (Extended Data Fig. 7a and Supplementary Table 23); the  $\beta$ -cell mass thus reached between 30% to 69% of the normal values, and remained stable for life (see earlier; Supplementary Table 6).

We characterized the  $\delta$ -cell-derived insulin $^+$  cells at the gene expression level by quantitative polymerase chain reaction (qPCR). We first compared islets isolated 2 weeks after  $\beta$ -cell ablation or after recovery (4 months post-DT) with age-matched control islets. Expression of all the  $\beta$ -cell-specific markers tested was robustly increased in recovered mice (Extended Data Fig. 7b). We also compared regenerated insulin $^+$  cells with native  $\beta$ -cells using sorted mCherry $^+$  cells obtained from either recovered or unablated age-matched (4.5-month-old) *insulin-mCherry; RIP-DTR* mice (Extended Data Fig. 7c). The two cell populations were very similar (Extended Data Fig. 7d), yet the  $\delta$ -cell-derived replicating  $\beta$ -cells displayed a potent downregulation of cyclin-dependent kinase





**Figure 2** |  $\delta$ -cells dedifferentiate, proliferate and reprogram into insulin production after extreme  $\beta$ -cell loss in *Sst-Cre; R26-YFP; RIP-DTR* juvenile mice. **a**, Immunofluorescence for YFP and Ki67 at 0.5 mpa. **b**, Eighty per cent of *Sst*-traced YFP<sup>+</sup> cells are Ki67<sup>+</sup> after  $\beta$ -cell ablation (controls:  $n = 6$ ; 2,754 YFP<sup>+</sup>-cells scored; DT:  $n = 6$ ; 3,146 YFP<sup>+</sup>-cells scored;  $P < 0.0001$ , Welch's test;  $P = 0.0022$ , Mann-Whitney). Ctrl, control. **c**, **d**, At 1.5 mpa, 90% of insulin<sup>+</sup> cells co-express YFP (controls:  $n = 3$ ; 6,480 insulin<sup>+</sup>-cells scored; DT:  $n = 7$ ; 1,592 insulin<sup>+</sup>-cells scored;  $P < 0.0001$ , Welch's test;  $P = 0.0167$ , Mann-Whitney). Arrow indicates YFP<sup>+</sup>/Sst<sup>+</sup> cells; arrowhead indicates YFP<sup>+</sup>/insulin<sup>+</sup> cells. **e**, In controls, 99.9% of the YFP<sup>+</sup> cells are Sst<sup>+</sup> ( $n = 3$ ; 1,673 YFP<sup>+</sup>-cells scored). In contrast, at 1.5 mpa only 55% of the YFP<sup>+</sup> cells are Sst<sup>+</sup>, while 45% of the YFP<sup>+</sup> cells are insulin<sup>+</sup> ( $n = 5$ ; 2,295 YFP<sup>+</sup>-cells scored;  $P < 0.0001$ , Welch's test;  $P = 0.0357$ , Mann-Whitney). **f**, Comparative evolution of glycaemia after  $\beta$ -cell ( $n = 5$ ),  $\delta$ -cell ( $n = 4$ ) and  $\beta$ - and  $\delta$ -cell co-ablation ( $n = 5$ ) in juveniles. **g**,  $\delta$ -cell conversion sequence. Scale bars, 20  $\mu$ m. Error bars show standard deviation (s.d.). \*\*\*\* $P < 0.0001$ .

inhibitors and regulators (Extended Data Fig. 7e, f). This suggests that reconstituted insulin<sup>+</sup> cells are like  $\beta$ -cells with transient proliferation capacity. Future studies will establish whether reconstituted ( $\delta$ ) $\beta$ -like cells are true equivalents to native  $\beta$ -cells.

qPCR and lineage-tracing analyses on islets isolated from pups at different regeneration time-points, together with *Ngn3* (also known as *Neurog3*) knockout induction after  $\beta$ -cell ablation, revealed that *Ngn3* transcription is required for the  $\delta$ -to-insulin<sup>+</sup> cell conversion to occur (Extended Data Fig. 8a–k and Supplementary Tables 24–29). Of note, the brief expression of *Ngn3* is a feature of islet precursor cells in the embryonic pancreas<sup>4</sup>. Together, these observations are compatible with a model in which  $\beta$ -cell reconstitution after ablation in juveniles occurs following a defined sequence of events:  $\delta$ -cells dedifferentiate, replicate once, and then half of the progeny activates *Ngn3* expression before insulin production (Fig. 2g). This was tested in a combined double lineage-tracing experiment using *Sst-Cre; R26-Tomato; Ngn3-YFP; RIP-DTR* mice. Six

weeks after  $\beta$ -cell ablation, insulin<sup>+</sup> cells in juveniles were Tomato<sup>+</sup>/YFP<sup>+</sup> (Extended Data Fig. 8k).

One key reprogramming and cell cycle entry player is FoxO1, a transcription factor whose downregulation triggers *Ngn3* expression in human fetal pancreatic explants<sup>5</sup> and favours insulin production in *Ngn3*<sup>+</sup> entero-endocrine progenitors<sup>6</sup>. FoxO1, usually in cooperation with TGF- $\beta$ /SMAD signalling<sup>7,8</sup>, inhibits cell proliferation through the transcriptional regulation of cell cycle inhibitors and activators<sup>9</sup>, and is involved in cellular senescence<sup>7</sup> (Extended Data Fig. 9a). We next explored the FoxO1 molecular network in purified adult or juvenile  $\delta$ -cells before and after (1 week)  $\beta$ -cell ablation, using *Sst-Cre; R26-YFP; RIP-DTR* mice.

$\delta$ -cells displayed divergent regulation of *Foxo1* in injured juvenile and adult mice. Consistent with *Foxo1* downregulation in juvenile  $\delta$ -cells, *Pdk1* and *Akt* (also known as *Akt2*) levels were increased, *Cdkn1a* (also known as *p21*) and *Cdkn2b* (also known as *p15Ink4b*) were downregulated, and *Cks1b*, *Cdk2* and *Skp* were upregulated (Fig. 3a), which is compatible with the proliferative capacity of juvenile  $\delta$ -cells after  $\beta$ -cell ablation. The opposite was found in the  $\delta$ -cells of ablated adults (Fig. 3a and Extended Data Fig. 9b).

Moreover, in  $\delta$ -cells of juveniles, but not in adults, there was a robust upregulation of BMP1/4 downstream effectors (Fig. 3b)<sup>10,11</sup>. Inversely, TGF- $\beta$  pathway genes were upregulated in  $\delta$ -cells of regenerating adults (Fig. 3b), which is compatible with the senescence scenario<sup>7</sup> involving PI3K/FoxO1 and TGF- $\beta$ /SMAD cooperation to maintain differentiation and cycle arrest (Extended Data Fig. 9a, b).

In summary, PI3K/AKT and SKP2/SCF pathways potentially cooperate to downregulate *Foxo1* in  $\delta$ -cells of regenerating juveniles. Also, upregulation of BMP effectors (*Id1* and *Id2*) could contribute to  $\delta$ -cell dedifferentiation and proliferation, as observed in other systems<sup>10,11</sup> (Fig. 3c). Conversely, the PI3K/AKT pathway remained downregulated in  $\delta$ -cells of ablated adults, which would allow FoxO1 to impede proliferation and dedifferentiation, probably through partnership with previously described SMADs<sup>12</sup> (Extended Data Fig. 9b).

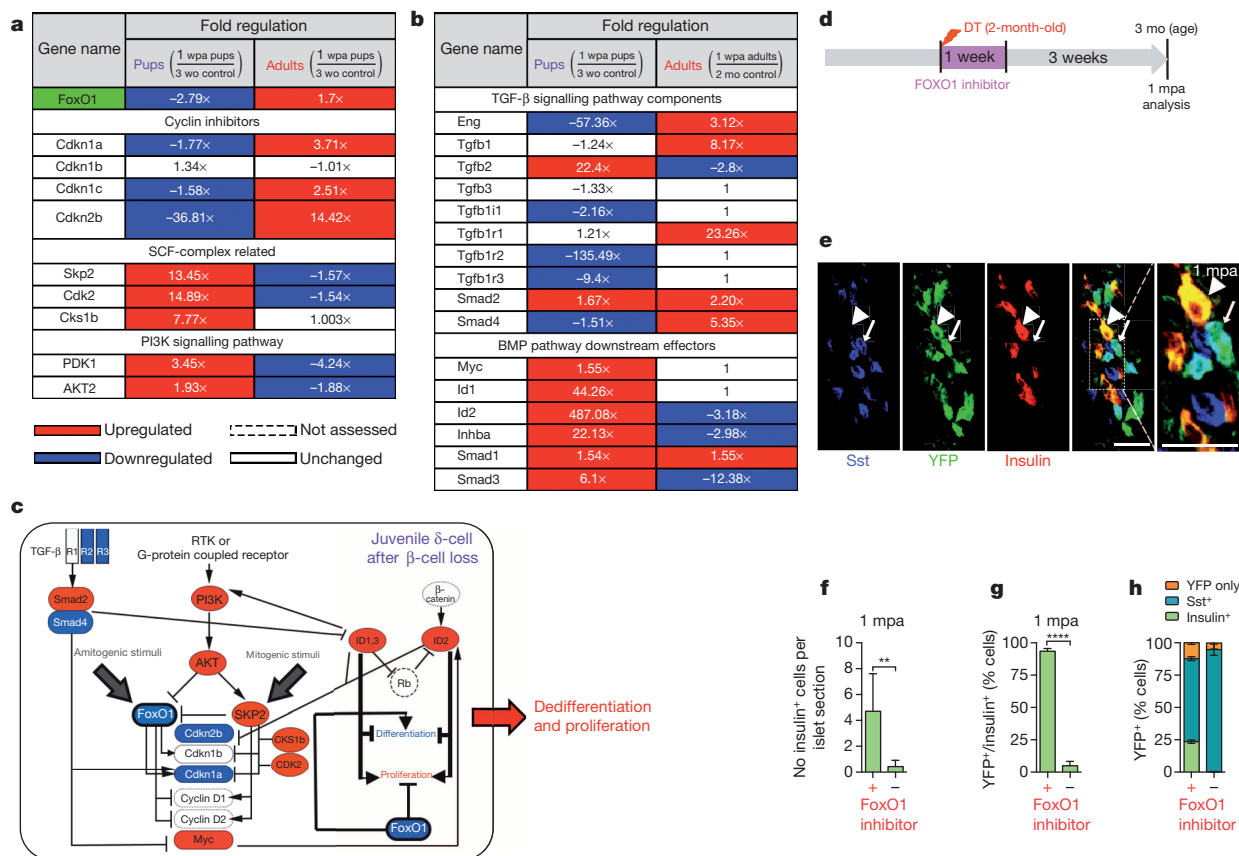
We next checked whether a transient FoxO1 inhibition in adult mice would lead to a juvenile-like  $\delta$ -to- $\beta$ -cell conversion. Indeed, inactivation of FoxO1 in  $\beta$ -cells causes their dedifferentiation<sup>13</sup>. Here, *Sst-Cre; R26-YFP; RIP-DTR*  $\beta$ -cell-ablated adult mice were given a FoxO1 inhibitor (AS1842856) for 1 week, either immediately following ablation (Fig. 3d) or 1 month later (Extended Data Fig. 10f and Supplementary Tables 37–39)<sup>14,15</sup>. While FoxO1 inhibition in non-ablated controls had a minimal effect on insulin expression (Extended Data Fig. 10a–d and Supplementary Tables 30–32), regeneration in diabetic mice was improved: insulin<sup>+</sup> cells were more abundant (11-fold; Fig. 3e, f and Supplementary Table 33), and were reprogrammed  $\delta$ -cells (93% were YFP<sup>+</sup>, Fig. 3g and Supplementary Table 34). One-fourth of the YFP<sup>+</sup> cells expressed insulin only (Fig. 3h, Extended Data Fig. 10e and Supplementary Tables 35, 36), revealing that, like in juveniles, an important fraction of  $\delta$ -cells had converted to insulin production.

These results support the involvement of a regenerative FoxO1 network and confirm that  $\delta$ -cell conversion can be pharmacologically induced in diabetic adults. FoxO1 blockade has a pleiotropic effect: inhibition of hepatic gluconeogenesis<sup>14,15</sup> and, as we have shown, promotion of  $\delta$ -cell reprogramming.

A century ago Morgan coined the terms 'epimorphosis' and 'morphallaxis' to designate, respectively, regeneration involving either cell dedifferentiation and proliferation or direct conversion from one cell type into another without proliferation<sup>16</sup>. Here we report in mammals an age-dependent switch ('adult transition') between epimorphic regeneration during youth, and a less efficient yet persistent throughout life proliferation-independent morphallactic mechanism.

Our findings uncover a novel role for  $\delta$ -cells; perhaps *Sst*<sup>+</sup> cells in the stomach, intestine or hypothalamus share the same capabilities. Intra-islet cell plasticity triggered by the disappearance of  $\beta$ -cells is influenced by age: the proliferation decline in ageing cells<sup>17</sup> would explain the need for an adult transition. Although less efficient,  $\alpha$ -cell plasticity remains long-time after  $\beta$ -cell loss since it is proliferation-independent.





**Figure 3 | Age-dependent effect of  $\beta$ -cell loss on  $\delta$ -cells.** **a, b**, Transcriptional variation of cell cycle regulators, PI3K/AKT/FoxO1 network genes (**a**), and TGF- $\beta$  and BMP components and effectors (**b**) in juvenile and adult  $\delta$ -cells 1 week after ablation, as compared with age-matched controls. wpa, weeks post-ablation. **c**,  $\beta$ -cell loss before puberty triggers FoxO1 downregulation in  $\delta$ -cells, while the opposite occurs in adults (see Extended Data Fig. 9b). **d**, Experimental design to transiently inhibit FoxO1 in  $\beta$ -cell-ablated adult mice. **e**, Induction of  $\delta$ -to-insulin cell conversion in diabetic adult mice. Dashed box indicates the area that is magnified in the right-most panel. Arrowheads indicate a converted  $\delta$ -cell, which has lost Sst expression (insulin<sup>+</sup> and YFP<sup>+</sup> cell). Arrows point to an unaffected  $\delta$ -cell, which is Sst<sup>+</sup> and YFP<sup>+</sup>, and does

These phenomena might be translatable to humans, as there is efficient  $\beta$ -cell regeneration in children with type 1 diabetes or after pancreatectomy<sup>9,18,19</sup>, and glucagon/insulin bihormonal human cells have been described upon epigenetic manipulation *ex vivo*<sup>20</sup>, and in diabetic patients<sup>21,22</sup>. Knowing also that only a small fraction of the  $\alpha$ -cell population is sufficient to maintain glucagon signalling<sup>23</sup>, understanding the nature of the diverse forms of intra-islet cell conversion might provide new opportunities for fostering the formation of ( $\alpha$ ) $\beta$ -like and ( $\delta$ ) $\beta$ -like cells.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 15 July 2013; accepted 30 June 2014.

Published online 20 August 2014.

- Matveyenko, A. V. & Butler, P. C. Relationship between  $\beta$ -cell mass and diabetes onset. *Diabetes Obes. Metab.* **10** (suppl. 4), 23–31 (2008).
- Atkinson, M. A. The pathogenesis and natural history of type 1 diabetes. *Cold Spring Harb. Perspect. Med.* <http://dx.doi.org/10.1101/cshperspect.a007641> (2012).
- Thorel, F. *et al.* Conversion of adult pancreatic  $\alpha$ -cells to  $\beta$ -cells after extreme  $\beta$ -cell loss. *Nature* **464**, 1149–1154 (2010).
- Desgraz, R. & Herrera, P. L. Pancreatic neurogenin 3-expressing cells are unipotent islet precursors. *Development* **136**, 3567–3574 (2009).
- Al-Masri, M. *et al.* Effect of forkhead box O1 (FOXO1) on  $\beta$  cell development in the human fetal pancreas. *Diabetologia* **53**, 699–711 (2010).

not express insulin. Scale bars, 20  $\mu$ m. **f, g**, Insulin<sup>+</sup> cells are 11-fold more abundant in FoxO1 inhibitor-treated mice (treated:  $n = 190$  islets, 4 mice; untreated:  $n = 95$  islets, 3 mice (inter-islet  $P < 0.0001$ , inter-individual  $P = 0.0065$ , Welch's test;  $P < 0.0001$ , Mann–Whitney) (**f**), and they are YFP<sup>+</sup> (93%) (treated:  $n = 4$ ; 894 insulin<sup>+</sup>-cells scored; untreated:  $n = 6$ ; 370 insulin<sup>+</sup>-cells scored;  $P < 0.0001$ , Welch's test;  $P = 0.0095$ , Mann–Whitney) (**g**). **h**, One-fourth of  $\delta$ -(YFP<sup>+</sup>) cells in adult  $\beta$ -cell-ablated FoxO1-inhibited mice dedifferentiate and become insulin expressors (treated:  $n = 4$ ; 3,358 YFP<sup>+</sup>-cells scored; untreated:  $n = 6$ ; 2,559 YFP<sup>+</sup>-cells scored). Error bars show s.d. \*\* $P < 0.01$ ; \*\*\*\* $P < 0.0001$ .

- Talchai, C., Xuan, S., Kitamura, T., DePinho, R. A. & Accili, D. Generation of functional insulin-producing cells in the gut by Foxo1 ablation. *Nature Genet.* **44**, 406–412 (2012).
- Muñoz-Espín, D. *et al.* Programmed cell senescence during mammalian embryonic development. *Cell* **155**, 1104–1118 (2013).
- Seoane, J., Le, H. V., Shen, L., Anderson, S. A. & Massague, J. Integration of Smad and Forkhead pathways in the control of neuroepithelial and glioblastoma cell proliferation. *Cell* **117**, 211–223 (2004).
- Karges, B. *et al.* Immunological mechanisms associated with long-term remission of human type 1 diabetes. *Diabetes Metab. Res. Rev.* **22**, 184–189 (2006).
- Yokota, Y. Id and development. *Oncogene* **20**, 8290–8298 (2001).
- Perk, J., Iavarone, A. & Benezra, R. Id family of helix-loop-helix proteins in cancer. *Nature Rev. Cancer* **5**, 603–614 (2005).
- van der Vos, K. E. & Coffey, P. J. FOXO-binding partners: it takes two to tango. *Oncogene* **27**, 2289–2299 (2008).
- Talchai, C., Xuan, S., Lin, H. V., Sussel, L. & Accili, D. Pancreatic  $\beta$  cell dedifferentiation as a mechanism of diabetic  $\beta$  cell failure. *Cell* **150**, 1223–1234 (2012).
- Nagashima, T. *et al.* Discovery of novel forkhead box O1 inhibitors for treating type 2 diabetes: improvement of fasting glycemia in diabetic db/db mice. *Mol. Pharmacol.* **78**, 961–970 (2010).
- Tanaka, H. *et al.* Effects of the novel Foxo1 inhibitor AS1708727 on plasma glucose and triglyceride levels in diabetic db/db mice. *Eur. J. Pharmacol.* **645**, 185–191 (2010).
- Morgan, T. H. Regeneration and liability to injury. *Science* **14**, 235–248 (1901).
- Chen, H. *et al.* PDGF signalling controls age-dependent proliferation in pancreatic  $\beta$ -cells. *Nature* **478**, 349–355 (2011).
- Karges, B. *et al.* Complete long-term recovery of  $\beta$ -cell function in autoimmune type 1 diabetes after insulin treatment. *Diabetes Care* **27**, 1207–1208 (2004).

19. Desgraz, R., Bonal, C. & Herrera, P. L.  $\beta$ -Cell regeneration: the pancreatic intrinsic faculty. *Trends Endocrinol. Metab.* **22**, 34–43 (2011).
20. Bramswig, N. C. *et al.* Epigenomic plasticity enables human pancreatic alpha to beta cell reprogramming. *J. Clin. Invest.* **123**, 1275–1284 (2013).
21. Butler, A. E. *et al.* Marked expansion of exocrine and endocrine pancreas with incretin therapy in humans with increased exocrine pancreas dysplasia and the potential for glucagon-producing neuroendocrine tumors. *Diabetes* **62**, 2595–2604 (2013).
22. Yoneda, S. *et al.* Predominance of  $\beta$ -cell neogenesis rather than replication in humans with an impaired glucose tolerance and newly diagnosed diabetes. *J. Clin. Endocrinol. Metab.* **98**, 2053–2061 (2013).
23. Thorel, F. *et al.* Normal glucagon signaling and  $\beta$ -cell function after near-total  $\alpha$ -cell ablation in adult mice. *Diabetes* **60**, 2872–2882 (2011).

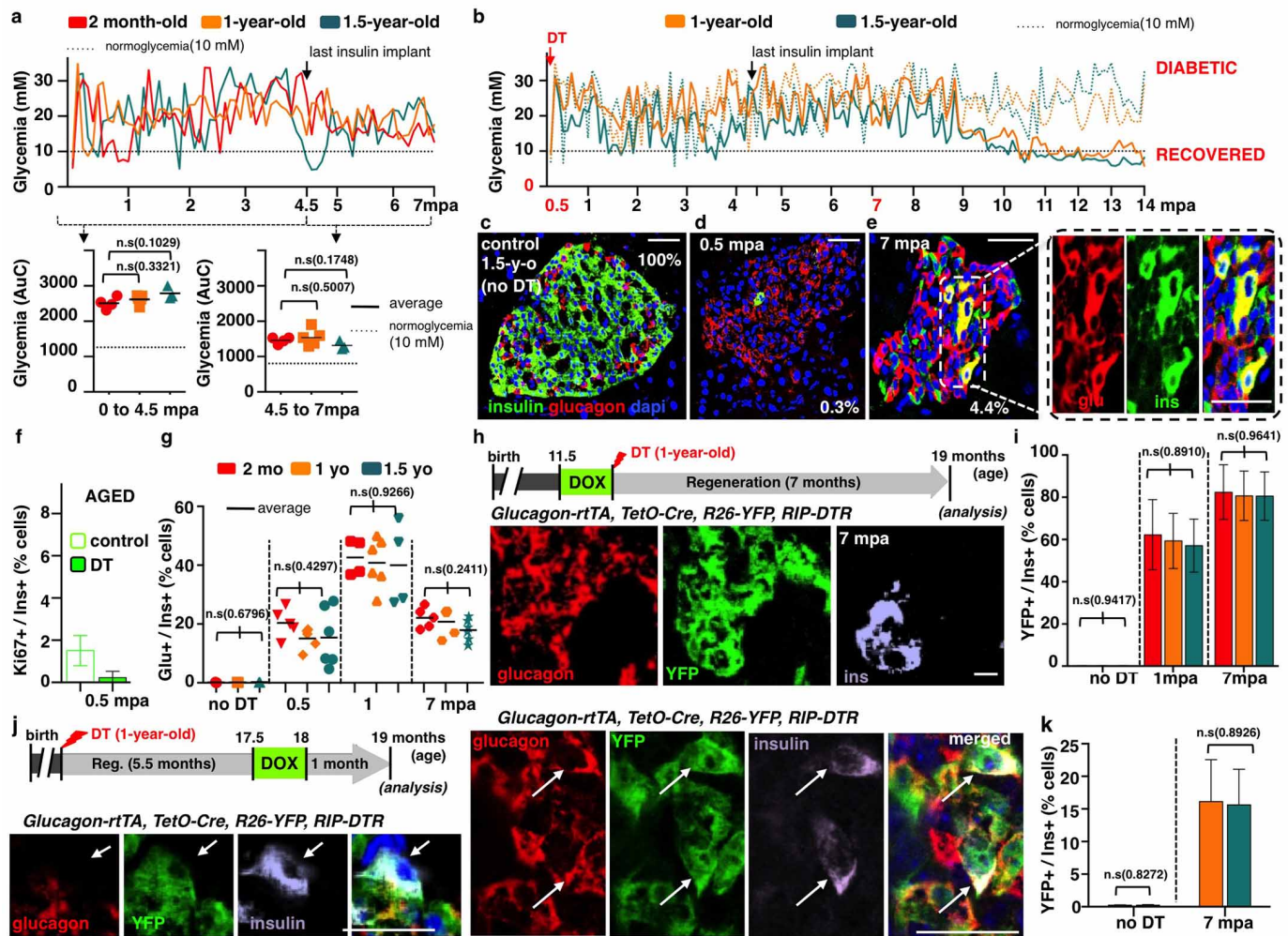
**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We are grateful to D. Belin, P. Vassalli, R. Stein, A. Cookson, A. Ruiz i Altaba, M. González Gaitán, B. Galliot and I. Rodríguez for comments, support and discussions, and to G. Gallardo, O. Fazio, K. Hammad and B. Polat for technical help. We

thank G. Gradwohl for the *Ngn3-YFP* mice. F.M.G. and F.R. were funded by Wellcome Trust grants WT088357/Z/09/Z and WT084210/Z/07/Z, respectively. This work was funded with grants from the National Institutes of Health/National Institute of Diabetes and Digestive and Kidney Diseases (Beta Cell Biology Consortium), the Juvenile Diabetes Research Foundation and the Swiss National Science Foundation (NRP63) to P.L.H.

**Author Contributions** S.C. conceived and performed the experiments and analyses, and wrote the manuscript. F.M.G. and F.R. generated the *Sst-Cre* line, and G.G. and J.N.J. generated the *Ngn3-CreERT*, *Ngn3-tTA* and *TRE-Ngn3* lines. D.B. characterized the pancreatic expression of the *Sst-Cre* line and performed the adult analysis. L.G. performed experiments and analyses. V.C. profiled sorted fluorescent adult islet cells. K.F. and F.T. performed immunofluorescence microscopy. P.L.H. conceived the experiments and wrote the manuscript.

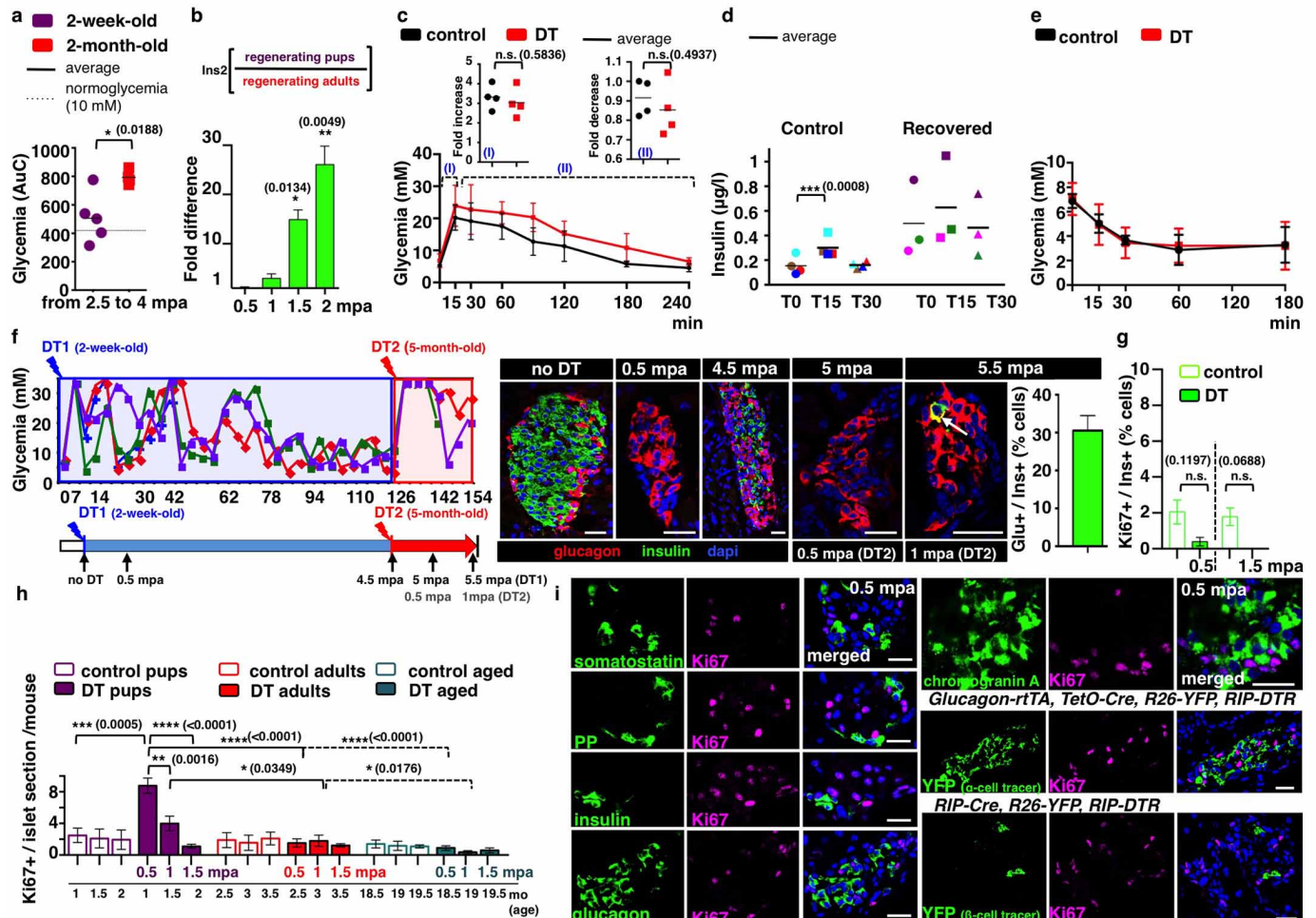
**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.H. ([pedro.herrera@unige.ch](mailto:pedro.herrera@unige.ch)).



**Extended Data Figure 1 | Maintenance of α-cell plasticity in diabetic aged mice.** **a**, Evolution of glycaemia in β-cell-ablated adults (middle-aged) and aged mice. The area under the curve (AuC) in middle-aged (2-month-old,  $n = 4$ ) and aged (1- and 1.5-year-old,  $n = 5$  and  $n = 3$ ) mice before and after stopping insulin administration revealed no statistical difference between groups (Welch's test,  $P_{0-4.5 \text{ mpa}} = 0.1029, 0.3321$ ;  $P_{4.5-7 \text{ mpa}} = 0.1748, 0.5007$ ; one-way analysis of variance (ANOVA),  $P = 0.1161, P = 0.2681$ ; and Mann-Whitney,  $P = 0.1640, 0.4519$ ). **b**, Evolution of glycaemia in 14 aged mice over a period of 14 months post-ablation (mpa). Mice were treated with insulin for 4.5 months; most of them (5/7 in each group) subsequently recovered from diabetes. **c-e**, Pancreatic islets before (c) and after (d, e) β-cell ablation in 1.5-year-old mice; β-cell mass increases 3.5-fold between 0.5 and 1 mpa, 12-fold at 7 mpa and 32-fold at 14 mpa, in all age groups. Percentages (0.3% and 4.4%) indicate β-cell mass relative to unablated controls (Supplementary Table 1). Two-month-old:  $n_{0.5 \text{ mpa}} = 4$ ;  $n_{1 \text{ mpa}} = 4$ ;  $n_{7 \text{ mpa}} = 4$ ; 1-year-old:  $n_{0.5 \text{ mpa}} = 5$ ,  $n_{1 \text{ mpa}} = 5$ ,  $n_{7 \text{ mpa}} = 5$ ,  $n_{14 \text{ mpa}} = 8$ ; 1.5-year-old:  $n_{0.5 \text{ mpa}} = 3$ ;  $n_{1 \text{ mpa}} = 3$ ;  $n_{7 \text{ mpa}} = 3$ ,  $n_{14 \text{ mpa}} = 8$ . **f**, β-Cell proliferation is very low in aged mice, whether control (1.5%;  $n = 8$ ; 39,790 insulin<sup>+</sup>-cells scored) or ablated (0.2%;  $n = 6$ ; 938 insulin<sup>+</sup>-cells scored) (Supplementary Table 2). **g**, Proportion of insulin<sup>+</sup> cells also containing glucagon after DT is not different between groups (Supplementary Table 3). Control:  $n_{2\text{-month-old}} = 3$ ;  $n_{1\text{-year-old}} = 3$ ;  $n_{1.5\text{-year-old}} = 3$ ; 0.5 mpa:  $n_{2\text{-month-old}} = 5$ ;  $n_{1\text{-year-old}} = 5$ ;  $n_{1.5\text{-year-old}} = 6$ ; 1 mpa:  $n_{2\text{-month-old}} = 5$ ;  $n_{1\text{-year-old}} = 6$ ;  $n_{1.5\text{-year-old}} = 4$ ; 7 mpa:  $n_{2\text{-month-old}} = 5$ ;  $n_{1\text{-year-old}} = 5$ ;  $n_{1.5\text{-year-old}} = 6$ . One-way ANOVA ( $P = 0.6796, 0.4297, 0.9266$ ,

0.2411); note that 40% of the cells containing insulin at 1 mpa also contained glucagon. The proportion of glucagon<sup>+</sup>/insulin<sup>+</sup> cells remains constant between 0.5 and 7 mpa, while the number of insulin<sup>+</sup> cells increases with time (e; Supplementary Table 1), suggesting that there is a cumulative recruitment of α-cells into insulin production. **h**, Islet with YFP<sup>+</sup>/glucagon<sup>+</sup>/insulin<sup>+</sup> cells in 1-year-old *glucagon-rtTA; TetO-Cre; R26-YFP; RIP-DTR* mice, 7 mpa; rtTA expression allows the selective irreversible YFP labelling of adult α-cells upon administration of doxycycline (DOX) before β-cell ablation. **i**, Proportion of YFP-labelled insulin-expressing cells in DOX-treated mice. Eighty per cent of insulin<sup>+</sup> cells are YFP<sup>+</sup> after 7 mpa, in all age groups (Supplementary Table 4). Control:  $n_{2\text{-month-old}} = 3$ ;  $n_{1\text{-year-old}} = 3$ ;  $n_{1.5\text{-year-old}} = 3$ ; 1 mpa:  $n_{2\text{-month-old}} = 5$ ;  $n_{1\text{-year-old}} = 3$ ;  $n_{1.5\text{-year-old}} = 3$ ; 7 mpa:  $n_{2\text{-month-old}} = 5$ ;  $n_{1\text{-year-old}} = 5$ ;  $n_{1.5\text{-year-old}} = 5$ . One-way ANOVA ( $P = 0.9417, 0.8910, 0.9641$ ). **j**, **k**, YFP<sup>+</sup>/glucagon<sup>+</sup>/insulin<sup>+</sup> cells at 7 mpa, following DOX pulse-labelling at 5.5 months after β-cell loss (Supplementary Table 5). Control:  $n_{1\text{-year-old}} = 5$ ;  $n_{1.5\text{-year-old}} = 5$ ; 7 mpa:  $n_{1\text{-year-old}} = 5$ ;  $n_{1.5\text{-year-old}} = 5$ ; Welch's correction ( $P = 0.8272, 0.8926$ ), Mann-Whitney ( $P = 0.9444$ ). On average, 15% of the insulin<sup>+</sup> cells found were YFP labelled, some of which no longer contained glucagon as in j, bottom row. Note the decreased proportion of YFP-labelled insulin<sup>+</sup> cells when α-cells are tagged late after ablation (from 80% to 15%; compare i and k), and the presence of YFP-labelled insulin<sup>+</sup>/glucagon-negative cells in the latter situation (j), suggesting that bihormonal α-cells slowly but gradually lose glucagon gene activity. Scale bars, 20 μm. Error bars show s.d.



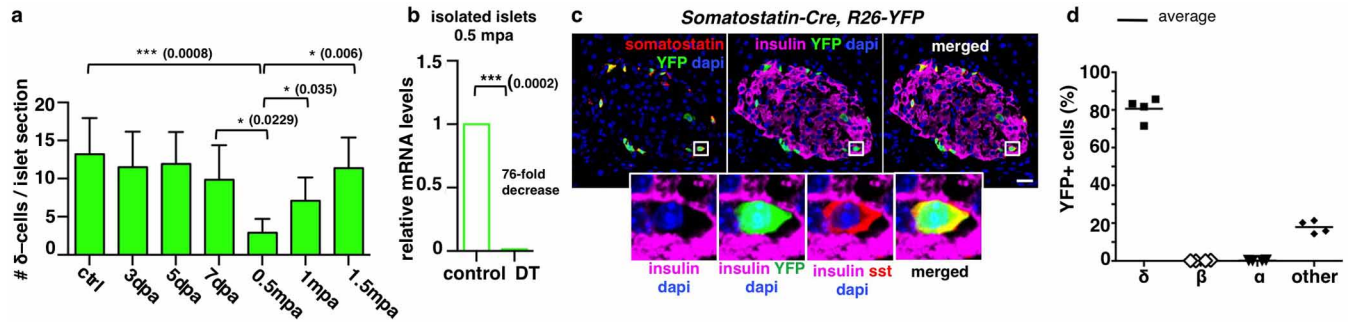


### Extended Data Figure 2 | Diabetes recovery in pre-pubertal mice.

**a**, Evolution of glycemia (AuC) between 2.5 and 4 mpa, in pups and adults (see Fig. 1b) (Welch's test,  $P = 0.0188$ ). **b**, qPCR of insulin 2 messenger RNA after  $\beta$ -cell ablation; insulin 2 transcripts are 25-fold more abundant in pups than in adults at 2 mpa ( $n = 3$  mice per group, each individual sample was run in triplicate in each reaction for a total of three independent reactions). Built-in Welch's test ( $P = 0.0134$ ,  $0.0049$ ). Error bars show s.d. **c**, Glucose tolerance tests (IPGTT) for DT-treated (4.5 mpa,  $n = 4$ ) and age-matched controls ( $n = 4$ ); note the fold increase between glucose injection and the glycaemic peak during IPGTT for each animal, and fold decrease between glycaemic peak and T120 (two-tailed unpaired  $t$ -test,  $P_I = 0.5836$ ,  $P_{II} = 0.4937$ ). **d**, Plasma insulin at time point (in min) T0, T15 and T30 during the IPGTT. Control:  $n = 4$ ; DT:  $n = 4$ ; two-tailed paired  $t$ -test ( $P = 0.0008$ ). **e**, Insulin tolerance tests (ITT) performed 1.5 years after  $\beta$ -cell ablation at 2 weeks of age. Controls:  $n = 7$ ; DT:  $n = 10$ . **f**, 4.5 months after  $\beta$ -cell ablation (at 2 weeks), three mice became normoglycaemic and received a second treatment with DT. Ablation of regenerated insulin<sup>+</sup> cells in recovered mice leads to the appearance of glucagon<sup>+</sup>/insulin<sup>+</sup> cells, corresponding to the type of 'α-cell-dependent' regeneration observed in adults (31% of insulin<sup>+</sup> cells also contained glucagon; Supplementary Table 8). Arrow indicates glucagon<sup>+</sup>/insulin<sup>+</sup> bihormonal cell.

Error bars show standard error of the mean (s.e.m.). **g**,  $\beta$ -cell proliferation is very low in regenerating pups (Supplementary Table 9). Control:  $n_{1\text{-month-old}} = 3$ , 6,006 insulin<sup>+</sup>-cells scored;  $n_{2\text{-month-old}} = 3$ , 6,358 insulin<sup>+</sup>-cells scored; DT:  $n_{0.5\text{ mpa}} = 5$ , 412 insulin<sup>+</sup>-cells scored;  $n_{1.5\text{ mpa}} = 3$ , 675 insulin<sup>+</sup>-cells scored; Welch's test ( $P = 0.1197$ ,  $P = 0.0688$ ). Error bars show s.e.m. **h**, Islet cell proliferation is increased (3.5-fold; Ki67<sup>+</sup> cells) in islets of DT-treated pups at 0.5 mpa. Control:  $n_{1\text{-month-old}} = 3$ , 95 islets scored;  $n_{1.5\text{-month-old}} = 3$ , 94 islets scored;  $n_{2\text{-month-old}} = 3$ , 90 islets scored;  $n_{2.5\text{-month-old}} = 3$ , 89 islets scored;  $n_{3\text{-month-old}} = 3$ , 91 islets scored;  $n_{3.5\text{-month-old}} = 3$ , 93 islets scored;  $n_{18.5\text{-month-old}} = 3$ , 83 islets scored;  $n_{19\text{-month-old}} = 3$ , 83 islets scored;  $n_{19.5\text{-month-old}} = 3$ , 88 islets scored; DT (2-week-old):  $n_{0.5\text{ mpa}} = 6$ , 333 islets scored;  $n_{1\text{ mpa}} = 3$ , 91 islets scored;  $n_{1.5\text{ mpa}} = 3$ , 90 islets scored; DT (2-month-old):  $n_{0.5\text{ mpa}} = 3$ , 76 islets scored;  $n_{1\text{ mpa}} = 3$ , 77 islets scored;  $n_{1.5\text{ mpa}} = 3$ , 81 islets scored; DT (1.5-year-old):  $n_{0.5\text{ mpa}} = 3$ , 74 islets scored;  $n_{1\text{ mpa}} = 3$ , 81 islets scored;  $n_{1.5\text{ mpa}} = 3$ , 77 islets scored. Error bars show s.d. Welch's test, one-way ANOVA ( $P < 0.001$ ), Mann-Whitney ( $P = 0.0238$ ). **i**, Ki67<sup>+</sup> cells are hormone, chromogranin-A-negative; lineage-traced α- and DT-spared β-cells are Ki67-negative. Scale bars, 20 μm.

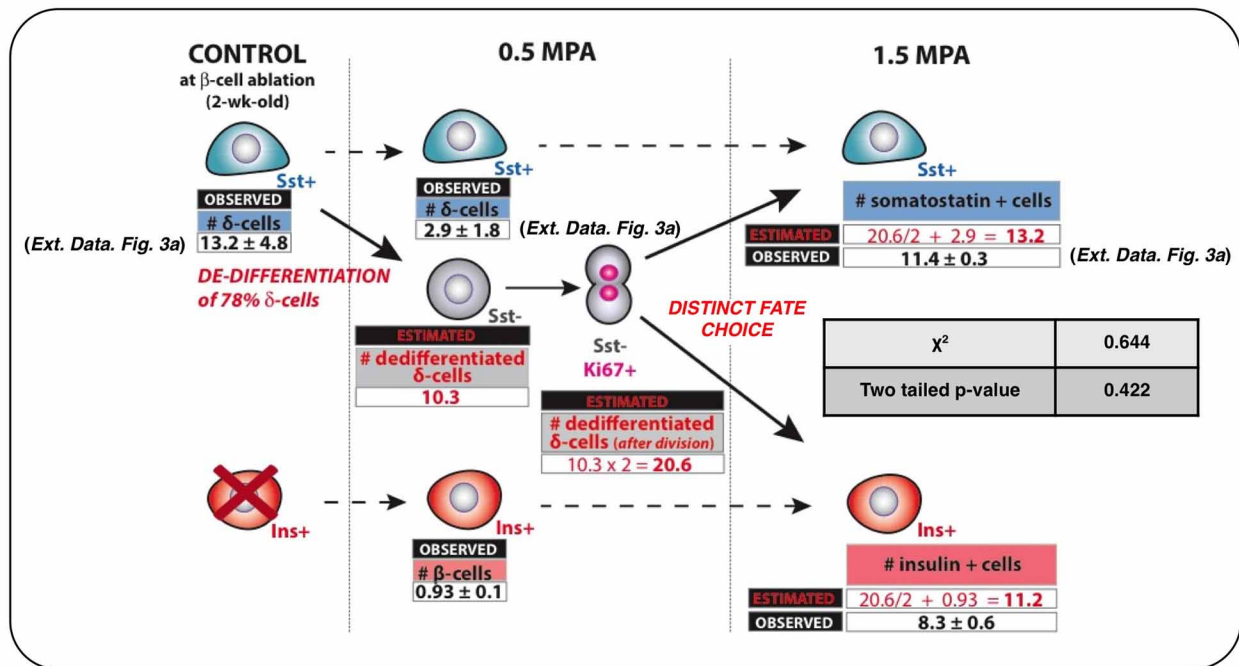




### Extended Data Figure 3 | $\delta$ -cell labelling and tracing in transgenic mice.

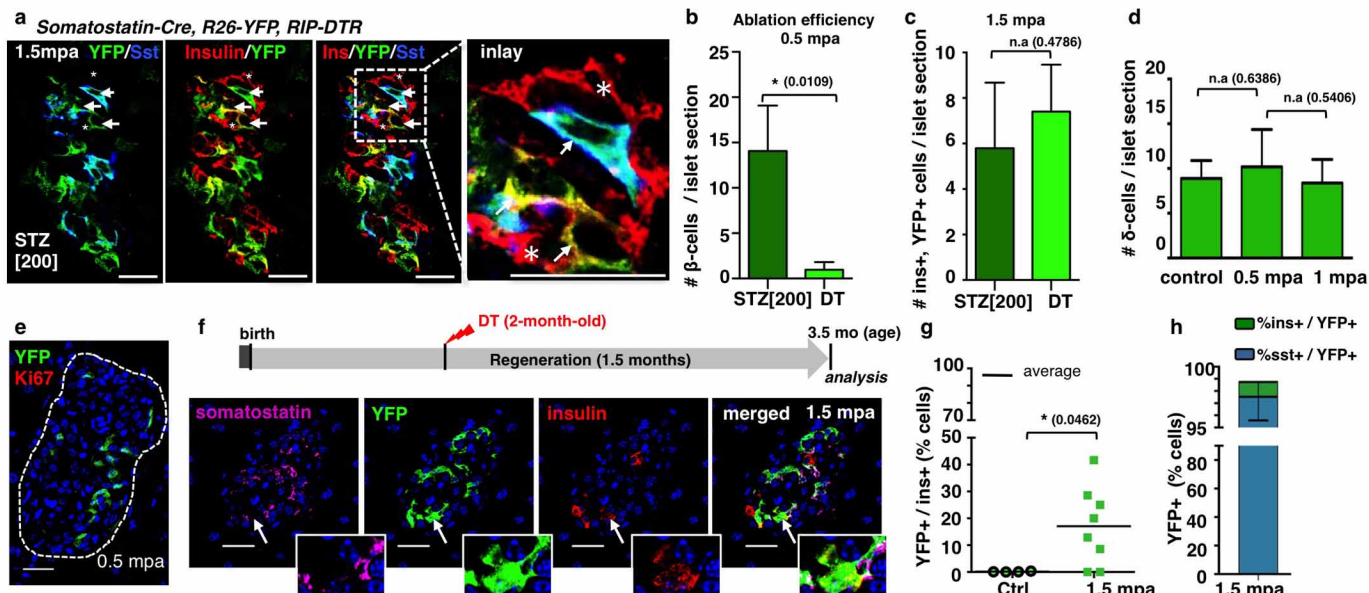
**a**, The number of  $Sst^+$  cells transiently decreases by 80% during the second week after ablation.  $n_{\text{control}} = 255$  islets, 7 mice;  $n_{3\text{dpa}} = 240$  islets, 5 mice;  $n_{5\text{dpa}} = 228$  islets, 5 mice;  $n_{7\text{dpa}} = 251$  islets, 5 mice;  $n_{0.5\text{mpa}} = 267$  islets, 6 mice;  $n_{1\text{mpa}} = 266$  islets, 5 mice;  $n_{1.5\text{mpa}} = 206$  islets, 5 mice. Error bars show s.d. Welch's test ( $P = 0.0008, 0.0229, 0.006, 0.035$ ), one-way ANOVA ( $P < 0.0001$ ), Mann-Whitney ( $P = 0.0043$ ). **b**, Relative *Sst* gene expression

sharply decreases 2 weeks after  $\beta$ -cell ablation in 2-week-old mice ( $n = 3$  mice per group, each individual sample of each experimental group was run in triplicate, in three independent reactions). Built-in Welch's test ( $P = 0.0002$ ). Error bars show s.d. **c**, *Sst-Cre; R26-YFP* mice. Cre activity efficiently and specifically occurs in  $\delta$ -cells (box: enlarged cell). Scale bar, 20  $\mu\text{m}$ . **d**, Quantitative values of reporter gene expression in islet cells ( $n = 4; 1,263$  YFP $^+$ -cells scored).



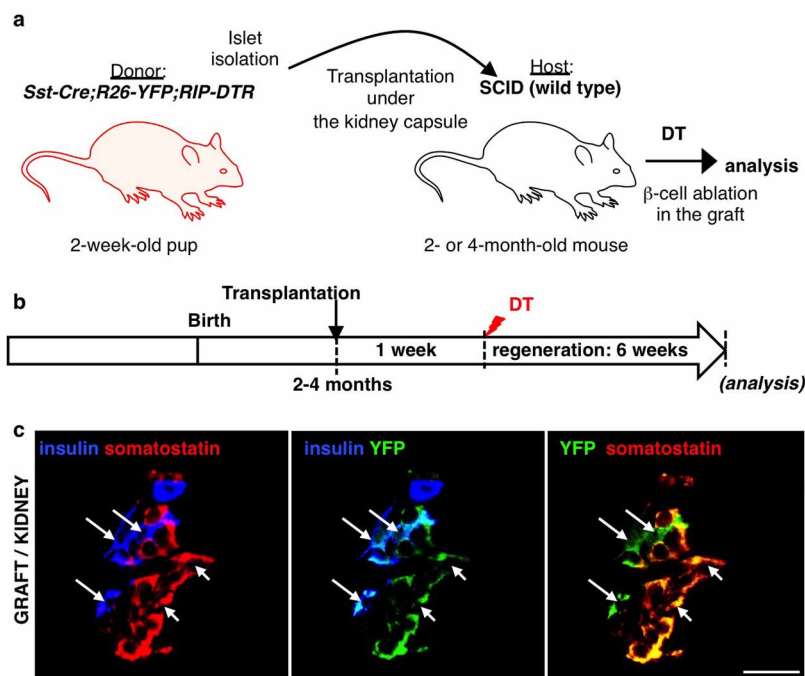
**Extended Data Figure 4 |  $\delta$ -cells dedifferentiate, proliferate and reprogram into insulin production after extreme  $\beta$ -cell loss in juvenile mice.** Observed and expected numbers of Sst<sup>+</sup> and insulin<sup>+</sup> cells per islet section, before and after  $\beta$ -cell ablation. Cells scored after 6 weeks (Extended Data Fig. 3a)

correspond ( $\chi^2$  test) with estimates made assuming that dedifferentiated proliferating  $\delta$ -cells yield two types of progeny (as deduced from Fig. 2c, e). Dashed arrows indicate phenotypic stability; plain arrows indicate dynamic behaviour (dedifferentiation and replication).



**Extended Data Figure 5 | Regeneration in streptozotocin-treated pups and DT-treated adults.** **a**, Immunofluorescence showing YFP-labelled insulin<sup>+</sup> cells at 1.5 month following streptozotocin (STZ)-induced ablation of  $\beta$ -cells in 2-week-old mice. Arrows indicate YFP<sup>+</sup>/insulin<sup>+</sup> cells; arrowhead indicates YFP<sup>+</sup>/Sst<sup>+</sup> cell; asterisks indicate escaping  $\beta$ -cells. **b**, Number of remaining  $\beta$ -cells per islet section at 2 weeks after streptozotocin or DT treatment in pups, reflecting difference in ablation efficiency of the two methods (Supplementary Table 18).  $n_{\text{STZ}} = 87$  islets, 3 mice;  $n_{\text{DT}} = 361$  islets, 4 mice. Welch's test (inter-islet  $P < 0.0001$ ; inter-individual  $P = 0.0109$ ), Mann-Whitney ( $P < 0.001$ ). **c**, The number of YFP<sup>+</sup>/insulin<sup>+</sup> cells per islet section at 1.5 mpa is not significantly different between the two  $\beta$ -cell ablation methods (Supplementary Table 19).  $n_{\text{STZ}} = 88$  islets, 3 mice;  $n_{\text{DT}} = 193$  islets, 7 mice. Welch's test ( $P = 0.4786$ ). **d**,  $\delta$ -cell numbers per islet section in controls

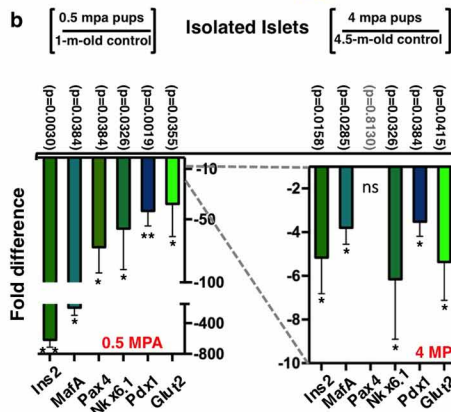
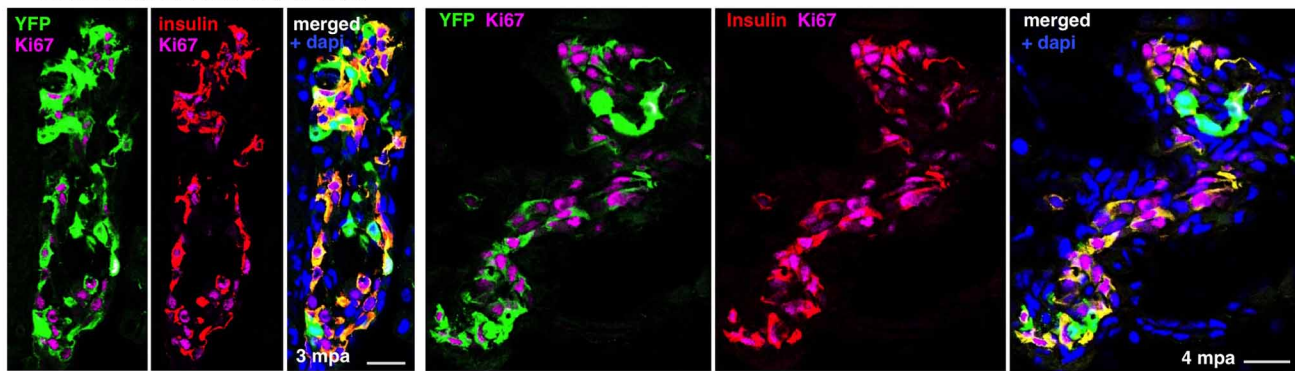
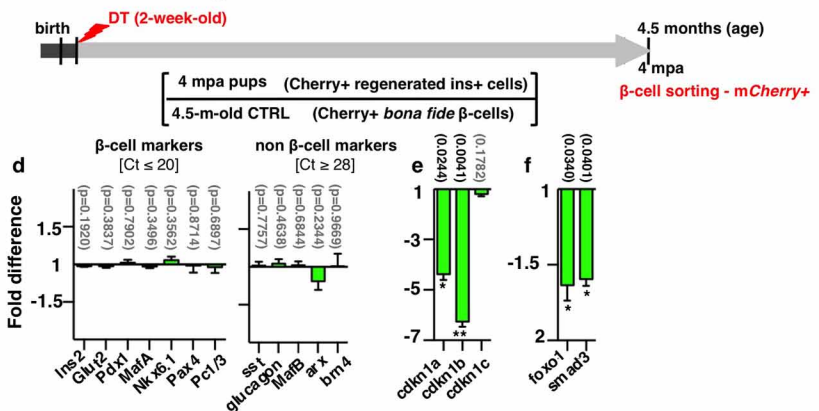
( $n = 3$ , 174 islets scored), 0.5 mpa ( $n = 4$ , 140 islets scored) and 1 mpa ( $n = 3$ , 86 islets scored). Unpaired  $t$ -test, two-tailed ( $P = 0.6386$ ;  $P = 0.5406$ ). **e**, Immunofluorescence for YFP and Ki67 2 weeks (0.5 mpa) after DT, in *Sst-Cre; R26-YFP; RIP-DTR* mice. **f**, Experimental design for  $\delta$ -cell tracing in  $\beta$ -cell-ablated *Sst-Cre; R26-YFP; RIP-DTR* mice at 2 months of age, and immunofluorescence for Sst, YFP and insulin at 1.5 mpa. Arrow indicates YFP<sup>+</sup>/insulin<sup>+</sup>/Sst<sup>+</sup> cell. **g**, At 1.5 mpa, 17% of insulin<sup>+</sup> cells co-express YFP versus almost 100% in ablated prepubescent mice. Control:  $n = 4$ ; DT:  $n = 8$ ; unpaired  $t$ -test, two-tailed ( $P = 0.0462$ ). **h**, At 1.5 mpa, 98% of the YFP<sup>+</sup> cells are Sst<sup>+</sup>, and 1% are insulin<sup>+</sup> cells (versus 44% in mice ablated before puberty;  $n = 8$ , unpaired  $t$ -test, two-tailed). Scale bars, 20  $\mu\text{m}$ . Error bars show s.d.



**Extended Data Figure 6 |  $\delta$ -to- $\beta$ -cell conversion after  $\beta$ -cell ablation is maintained in young islets ablated underneath the kidney capsule of adult hosts.** **a**, Islet transplantation design: 400–600 islets isolated from 2-week-old *Sst-Cre; R26-YFP; RIP-DTR* transgenics were transferred under the kidney capsule of 2-month-old immunodeficient (SCID) mice ( $n = 3$ ).

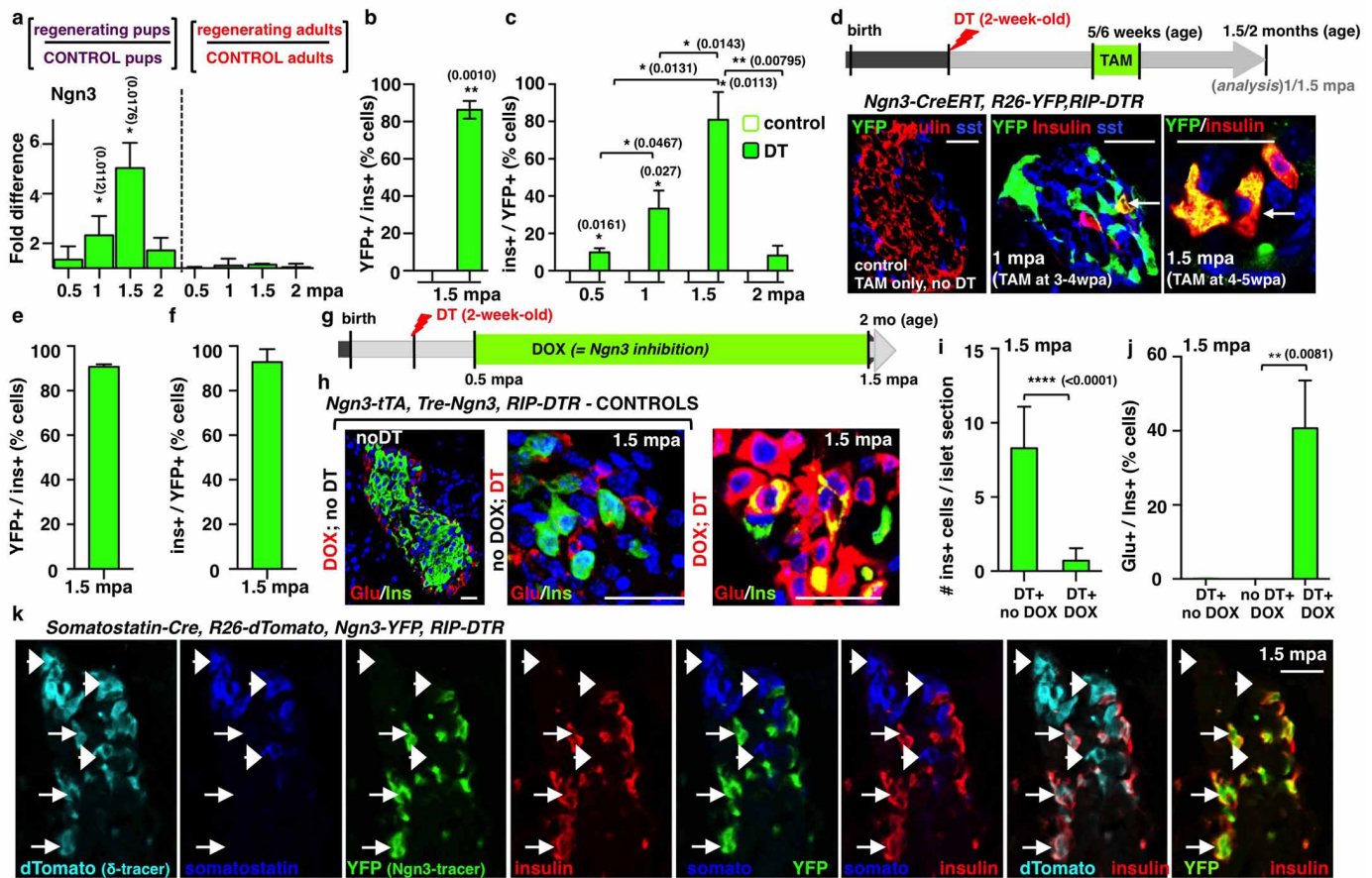
**b**, Experimental design: after 1 week of engraftment, adult host mice were DT-treated and left to regenerate for 6 weeks. **c**,  $\delta$ -to- $\beta$  conversion was observed in  $\beta$ -cell-ablated engrafted islets, like in the pancreas of juvenile mice. Scale bars, 20  $\mu$ m.



**a** *Somatostatin-Cre, R26-YFP, RIP-DTR***c** *Insulin-mCherry, RIP-DTR*

**Extended Data Figure 7 | Characterization of  $\delta$ -cell-derived regenerated insulin<sup>+</sup> cells.** **a**, Once differentiated from  $\delta$ -cells (YFP<sup>+</sup>), the newly formed  $\beta$ -cells re-enter the cell cycle (Ki67<sup>+</sup> cells). Two waves of massive replication occur, at 3 and 4 months after injury, respectively (Supplementary Table 23). **b**, qPCR for  $\beta$ -cell-specific genes using RNA extracted from islets isolated from control and DT-treated mice, either 2 weeks or 4 months after DT administration (0.5 mpa and 4 mpa). Note that after an initial extreme downregulation of all the  $\beta$ -cell-specific markers explored, their levels significantly recover after 4 months, which correlates with the observed robust regeneration and diabetes recovery. Values represent the ratio between each regeneration time-point and its age-matched control. **c**, Experimental design. **d**, qPCR comparison between regenerated mCherry<sup>+</sup>/insulin<sup>+</sup> cells isolated

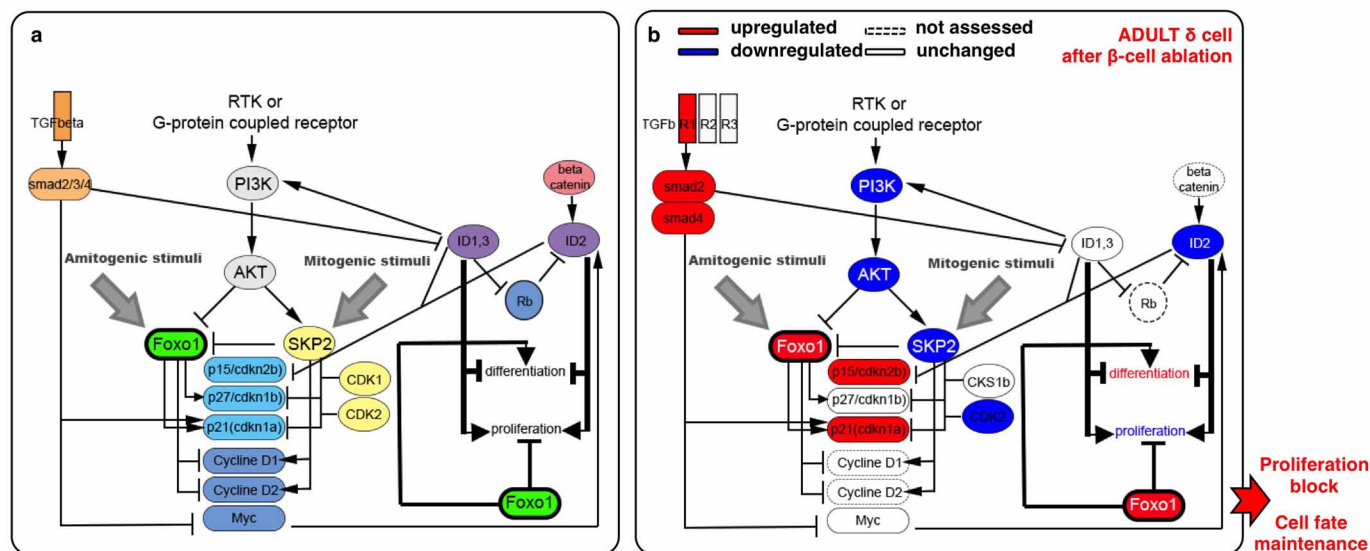
from mice 4 months after  $\beta$ -cell ablation, and mCherry<sup>+</sup>  $\beta$ -cells obtained from age-matched controls (4.5-month-old). All markers tested are expressed at identical levels in both groups; non- $\beta$ -cell markers are expressed at extremely reduced levels (threshold cycle (CT) ranging from 28 to 31), showing the same degree of purity in both types of cell preparations. **e**, **f**, Interestingly, in contrast to bona fide  $\beta$ -cells isolated from 4.5-month-old controls, regenerated insulin<sup>+</sup> cells have lower levels of cyclin-dependent kinase inhibitors, FoxO1 and Smad3. This correlates with their increased proliferative capacity at this specific time-point. Scale bars, 20  $\mu$ m. qPCRs:  $n = 3$  mice per group; each individual sample of each experimental group was run in triplicate, in three independent reactions; built-in Welch's test. Error bars show s.d.



### Extended Data Figure 8 | *Ngn3* activation is required for insulin expression in dedifferentiated $\delta$ -cells.

**a**, qPCR for *Ngn3* mRNA after  $\beta$ -cell ablation reveals a transitory fivefold upregulation of *Ngn3* transcripts 6 weeks after  $\beta$ -cell ablation when  $\beta$ -cell ablation is performed before puberty, but not in adult mice. Controls:  $n_{1\text{-month-old}} = 3$ ;  $n_{1.5\text{-month-old}} = 3$ ;  $n_{2\text{-month-old}} = 6$ ;  $n_{2.5\text{-month-old}} = 3$ ;  $n_{2.5\text{-month-old}} = 3$ ;  $n_{3\text{-month-old}} = 3$ ;  $n_{3.5\text{-month-old}} = 3$ ;  $n_{4\text{-month-old}} = 3$ ; DT (2-week-old):  $n_{0.5\text{ mpa}} = 3$ ;  $n_{1\text{ mpa}} = 3$ ;  $n_{1.5\text{ mpa}} = 6$ ;  $n_{2\text{ mpa}} = 3$ ; DT (2-month-old):  $n_{0.5\text{ mpa}} = 3$ ;  $n_{1\text{ mpa}} = 3$ ;  $n_{1.5\text{ mpa}} = 3$ ;  $n_{2\text{ mpa}} = 3$ . Each individual sample (mouse) was run in triplicate, in each of three independent reactions. Built-in Welch's test ( $P = 0.0112$ ,  $0.0178$ ). **b**, *Ngn3* transcriptional activity can be monitored in *Ngn3-YFP* knock-add-on mice because *Ngn3* promoter activity results in YFP expression. In non-ablated age-matched control pups, or in ablated adults, no islet YFP<sup>+</sup> cells were found (data not shown), yet when  $\beta$ -cells are ablated at 2 weeks of age, 86% of insulin<sup>+</sup> cells also express YFP<sup>+</sup> at 1.5 mpa. Control:  $n = 3$ , 6,358 insulin<sup>+</sup>-cells scored; DT:  $n = 3$ , 675 insulin<sup>+</sup>-cells scored; Welch's test ( $P = 0.0010$ ). **c**, At 1.5 mpa, 81% of YFP<sup>+</sup> cells co-express insulin, but no glucagon, Sst or PP (data not shown). Two weeks later, YFP<sup>+</sup> cells are almost absent, reflecting the downregulation of *Ngn3* expression reported in **a**, and suggesting that insulin<sup>+</sup> cells originate from cells transiently activating *Ngn3* expression after ablation. Control:  $n_{1\text{-month-old}} = 3$ ;  $n_{1.5\text{-month-old}} = 3$ ;  $n_{2\text{-month-old}} = 3$ ;  $n_{2.5\text{-month-old}} = 3$ ; absent YFP<sup>+</sup> cells in all control conditions; DT:  $n_{0.5\text{ mpa}} = 3$ , 31 YFP<sup>+</sup> cells;  $n_{1\text{ mpa}} = 3$ , 123 YFP<sup>+</sup> cells;  $n_{1.5\text{ mpa}} = 3$ , 729 YFP<sup>+</sup> cells;  $n_{2\text{ mpa}} = 3$ , 47 YFP<sup>+</sup> cells. Welch's test and ANOVA ( $P < 0.0001$ ). **d**, Irreversible lineage tracing of *Ngn3*-expressing cells at 1 and 1.5 mpa upon tamoxifen (TAM) administration in *Ngn3-CreERT; R26-YFP; RIP-DTR* mice; immunofluorescence analyses reveal that in the absence of  $\beta$ -cell ablation, there is no YFP induction (controls). In ablated mice, nearly all insulin<sup>+</sup> cells are YFP<sup>+</sup> with time (arrows). At early time-points (1 mpa), YFP<sup>+</sup>/hormone-negative cells are found: these are probably differentiating cells before insulin expression.

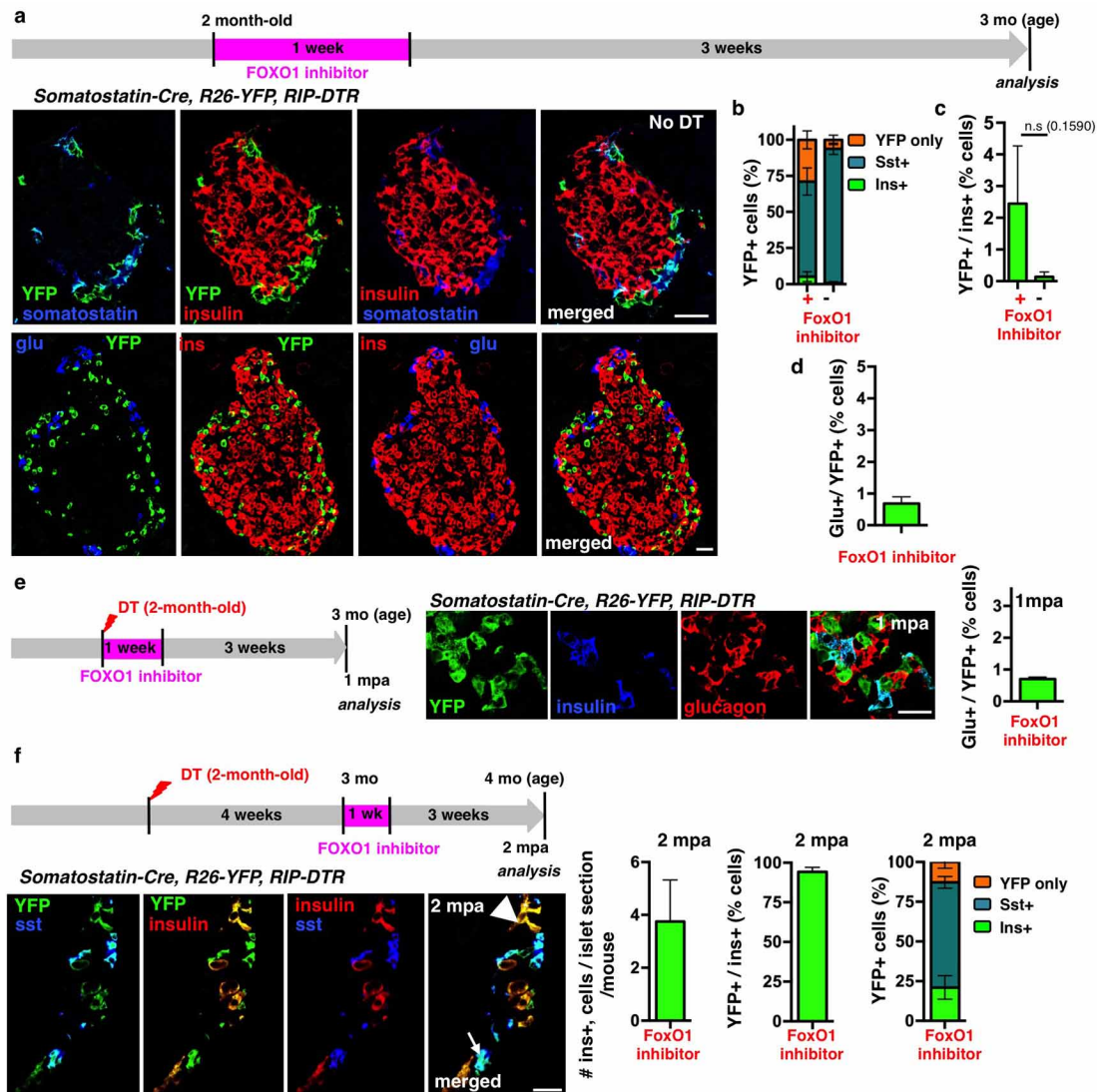
**e, f**, In  $\beta$ -cell-ablated *Ngn3-CreERT; R26-YFP; RIP-DTR* pups, 91% of insulin<sup>+</sup> cells co-express YFP<sup>+</sup> (control:  $n = 3$ , 3,472 insulin<sup>+</sup>-cells scored; DT:  $n = 3$ , 489 insulin<sup>+</sup>-cells scored) (**e**) and inversely, 93% of the YFP<sup>+</sup> cells are insulin<sup>+</sup> (**f**) (control:  $n = 3$ ; absent YFP<sup>+</sup>-cells in all control conditions; DT:  $n = 3$ , 478 YFP<sup>+</sup>-cells scored). **g**, Experimental design to block *Ngn3* upregulation in  $\beta$ -cell-ablated prepubescent mice by administering DOX to mice bearing five mutant alleles: *Ngn3-tTA*<sup>+/+</sup>; *TRE-Ngn3*<sup>+/+</sup>; *RIP-DTR*. In these mice the *Ngn3* coding region is replaced by a DOX-sensitive transactivator gene (tTA); the endocrine pancreas develops normally because *Ngn3* expression is allowed in the absence of DOX by the binding of tTA to the promoter of the *TRE-Ngn3* transgene. Pups were given DT at 2 weeks of age and then DOX 2 weeks later, to block *Ngn3* upregulation. They were euthanized when *Ngn3* peaks after ablation (2-month-old). **h**, Islets from non-ablated (no DT) and ablated (DT) mice, exposed (*Ngn3* inhibition) or not (normal *Ngn3* expression) to DOX treatment from 4 weeks of age.  $\beta$ -Cell regeneration is efficient in absence of DOX (as previously shown), but decreases after *Ngn3* blockade, resulting in the appearance of glucagon/insulin bihormonal cells. **i**, Sharply decreased regeneration by blocking *Ngn3* expression in DOX-treated mice reveals the requirement of *Ngn3* for efficient  $\beta$ -cell regeneration in pups. DT:  $n = 266$  islets scored, 3 mice; DT+DOX:  $n = 167$ , 4 mice. Welch's test (inter-islet  $P < 0.0001$ ; inter-animal  $P = 0.0352$ ), Mann-Whitney ( $P < 0.0001$ ). **j**, Glucagon<sup>+</sup>/insulin<sup>+</sup> bihormonal cells appear in DOX-treated  $\beta$ -cell-ablated pups (*Ngn3* inhibition), suggesting a switch to an 'adult-like', less efficient, mechanism of regeneration. Control+DOX:  $n = 3$ , 9,233 insulin<sup>+</sup>-cells scored; DT:  $n = 3$ , 1,385 insulin<sup>+</sup>-cells scored; DT+DOX:  $n = 4$ , 141 insulin<sup>+</sup>-cells scored. Welch's test ( $P = 0.0081$ ), ANOVA ( $P < 0.0001$ ). **k**, Combined double lineage tracing of  $\delta$ -cells (Tomato<sup>+</sup>) and *Ngn3*-expressing cells (YFP<sup>+</sup>) shows by immunofluorescence that nearly all insulin<sup>+</sup> cells express both reporters, but no Sst (arrows). Sst<sup>+</sup> cells (arrowheads) are YFP<sup>+</sup>- and insulin-negative. Scale bars, 20  $\mu\text{m}$ . Error bars show s.d.



**Extended Data Figure 9 | FoxO1 regulatory network.** **a**, Cartoon depicting the FoxO1 network involved in the regulation of cell cycle progression and cellular senescence: FoxO1 arrests the cell cycle by repressing activators (cyclin D1, cyclin D2) and inducing inhibitors (Cdkn1a, Cdkn1b, Cdkn2b, Cdkn1c) (PMID: 10102273; PMID: 17873901). Cdkn1a and Cdkn2b activation, a sign of cellular senescence (PMID: 17667954), is regulated by FoxO1 through direct interaction with Skp2 protein. In turn, Skp2 blocks FoxO1 and, together with CKS1b, CDK1 and CDK2, triggers the direct degradation of Cdkn1a and Cdkn1b, thus promoting proliferation (PMID: 15668399). FoxO proteins are inhibited mainly through PI3K/AKT-mediated phosphorylation (PMID: 10102273; PMID: 12621150; PMID: 21708191; PMID: 10217147; PMID:

17604717]: PDK1, the master kinase of the pathway, stimulates cell proliferation and survival by directly activating AKT, which phosphorylates (inhibits) the FoxOs (PMID: 10698680; PMID: 19635472). The PI3K/AKT/FoxO1 circuit requires active TGF- $\beta$ /SMAD signalling (PMID: 24238962; PMID: 15084259) in order to co-regulate Cdkn1a-dependent cell senescence. Active TGF- $\beta$  signalling downregulates the BMP pathway downstream effectors ID1 and ID2, known to promote dedifferentiation and proliferation during embryogenesis and cancer progression, probably through Cdkn2b regulation (PMID: 11840321; PMID: 16034366). **b**,  $\beta$ -cell ablation in adults triggers FoxO1 upregulation and the subsequent cell cycle arrest in  $\delta$ -cells.





**Extended Data Figure 10 |  $\delta$ -cell dedifferentiation in adult mice upon transient FoxO1 inhibition.** **a–d**, The 1 week FoxO1 inhibition with the compound AS1842856 in control unblated adult mice (**a**) results in dedifferentiation of one-fourth of the  $\delta$ -cell population (**b**; Supplementary Table 30) (treated:  $n = 3$ , 1,347 YFP<sup>+</sup>-cells scored; untreated:  $n = 4$ , 1,224 YFP<sup>+</sup>-cells scored; error bars show s.d.), without leading to insulin (**c**; Supplementary Table 31) (treated:  $n = 3$ , 3,249 insulin<sup>+</sup>-cells scored; untreated:  $n = 4$ , 9,562 insulin<sup>+</sup>-cells scored; error bars show s.d.; Welch's test ( $P = 0.1590$ )) or glucagon (**d**; Supplementary Table 32) (treated:  $n = 2$ , 728

YFP<sup>+</sup>-cells scored; error bars show s.e.m.) expression. **e**, One month following FoxO1 transient inhibition in  $\beta$ -cell-ablated adults, dedifferentiated  $\delta$ -cells do not express glucagon (Supplementary Table 36) (treated:  $n = 2$ , 986 YFP<sup>+</sup>-cells scored; error bars show s.e.m.). **f**, Transient FoxO1 inhibition a long time (1 month) after  $\beta$ -cell ablation also leads to the appearance of lineage-traced dedifferentiated  $\delta$ -cells that express insulin (Supplementary Tables 37–39) (treated:  $n = 3$ , 71 islets scored; 300 insulin<sup>+</sup>-cells scored; 1,216 YFP<sup>+</sup>-cells scored; error bars show s.d.). Scale bars, 20  $\mu$ m.



# High-fat-diet-mediated dysbiosis promotes intestinal carcinogenesis independently of obesity

Manon D. Schulz<sup>1\*</sup>, Çiğdem Atay<sup>1\*</sup>, Jessica Heringer<sup>1\*</sup>, Franziska K. Romrig<sup>1</sup>, Sarah Schwitalla<sup>1</sup>, Begüm Aydin<sup>2</sup>, Paul K. Ziegler<sup>3,4,5</sup>, Julia Varga<sup>3,4,5</sup>, Wolfgang Reindl<sup>6</sup>, Claudia Pommerenke<sup>7</sup>, Gabriela Salinas-Riester<sup>7</sup>, Andreas Böck<sup>8</sup>, Carl Alpert<sup>9†</sup>, Michael Blaut<sup>9</sup>, Sara C. Polson<sup>10</sup>, Lydia Brandl<sup>11</sup>, Thomas Kirchner<sup>4,5,11</sup>, Florian R. Greten<sup>3,4,5</sup>, Shawn W. Polson<sup>10</sup> & Melek C. Arkan<sup>1</sup>

Several features common to a Western lifestyle, including obesity and low levels of physical activity, are known risk factors for gastrointestinal cancers<sup>1</sup>. There is substantial evidence suggesting that diet markedly affects the composition of the intestinal microbiota<sup>2</sup>. Moreover, there is now unequivocal evidence linking dysbiosis to cancer development<sup>3</sup>. However, the mechanisms by which high-fat diet (HFD)-mediated changes in the microbial community affect the severity of tumorigenesis in the gut remain to be determined. Here we demonstrate that an HFD promotes tumour progression in the small intestine of genetically susceptible, *K-ras*<sup>G12Dint</sup> mice independently of obesity. HFD consumption, in conjunction with *K-ras* mutation, mediated a shift in the composition of the gut microbiota, and this shift was associated with a decrease in Paneth-cell-mediated antimicrobial host defence that compromised dendritic cell recruitment and MHC class II molecule presentation in the gut-associated lymphoid tissues. When butyrate was administered to HFD-fed *K-ras*<sup>G12Dint</sup> mice, dendritic cell recruitment in the gut-associated lymphoid tissues was normalized, and tumour progression was attenuated. Importantly, deficiency in MYD88, a signalling adaptor for pattern recognition receptors and Toll-like receptors, blocked tumour progression. The transfer of faecal samples from HFD-fed mice with intestinal tumours to healthy adult *K-ras*<sup>G12Dint</sup> mice was sufficient to transmit disease in the absence of an HFD. Furthermore, treatment with antibiotics completely blocked HFD-induced tumour progression, suggesting that distinct shifts in the microbiota have a pivotal role in aggravating disease. Collectively, these data underscore the importance of the reciprocal interaction between host and environmental factors in selecting a microbiota that favours carcinogenesis, and they suggest that tumorigenesis is transmissible among genetically predisposed individuals.

Undoubtedly, a variety of factors contribute to the aetiology of intestinal cancer. There are compelling arguments to include an HFD and the composition of the gut microbiota as key risk factors<sup>4</sup>. Given the rapid increase in the incidence of diet-induced obesity worldwide<sup>5</sup> and the recent evidence that the microbiota in obese individuals is more efficient at harvesting nutrients<sup>6,7</sup>, it is crucial to understand the role of the gut microbiota in the pathogenesis of cancer.

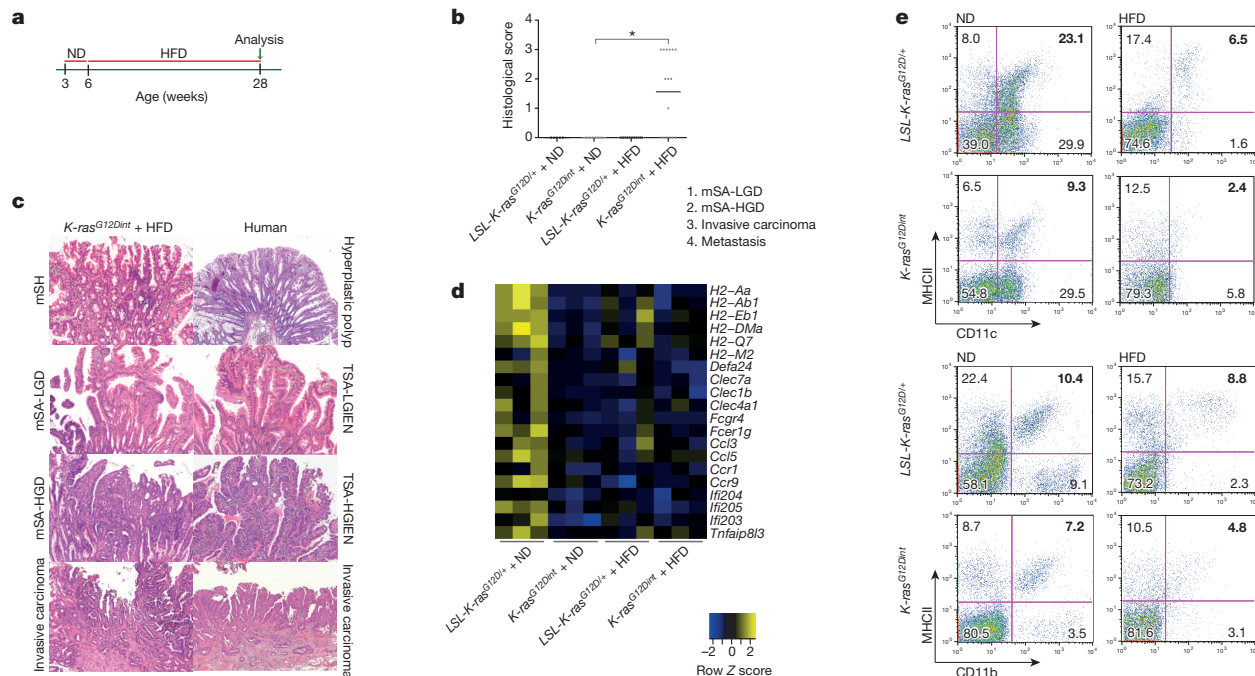
We set out to elucidate whether alterations in the gut microbial community are the link between an HFD and pathogenesis of intestinal cancer. To that end, we used a well-characterized serrated hyperplasia model with oncogenic *K-ras* expression in the intestinal epithelium (*K-ras*<sup>G12Dint</sup> mice)<sup>8</sup> and exposed these mice to an HFD regimen for 22 weeks (Fig. 1a). Whereas 33% of age-matched *K-ras*<sup>G12Dint</sup> mice on a normal diet developed only murine serrated hyperplasia along the

intestine, HFD consumption led to further tumour progression in 60% of *K-ras*<sup>G12Dint</sup> mice (Fig. 1b). These mice developed tumours in the duodenum that ranged from murine serrated adenoma, with low-grade dysplasia (mSA-LGD) and high-grade dysplasia (mSA-HGD), to invasive carcinoma, closely recapitulating the carcinogenic sequence in humans (Fig. 1c). These cells metastasized to the liver, pancreas and spleen when mice were maintained on an HFD for 40 weeks (Extended Data Fig. 1a). Serrated tumours revealed increased proliferation at the tips of the villi; such proliferation was otherwise restricted to the crypt region (Extended Data Fig. 1b). Importantly, diet-induced obesity was compromised in *K-ras*<sup>G12Dint</sup> mice during disease progression (Extended Data Fig. 1c). In accordance with the increased insulin sensitivity observed in these animals, lipid accumulation in the liver was lower than in littermate controls that were fed an HFD for 22 weeks (Extended Data Fig. 1d, e). Taken together, these data suggest that diet-mediated effects on the host were responsible for promoting serrated tumour progression in the small intestine.

An HFD induces a low-grade inflammatory state<sup>9</sup>, and inflammation is a hallmark of cancer<sup>10</sup>. Unexpectedly, tumour-necrosis factor- $\alpha$  (TNF- $\alpha$ ), as well as the macrophage cell surface marker F4/80, were downregulated in the duodenum of *K-ras*<sup>G12Dint</sup> mice (Extended Data Fig. 1f). Moreover, duodenal samples from *K-ras*<sup>G12Dint</sup> animals exhibited a significant reduction in the expression of genes involved in the recognition of, and immune response to, antigens, suggesting that an oncogene-driven downmodulation of host immunity had occurred (Fig. 1d and Extended Data Fig. 2a). This downregulation was associated with decreased Paneth-cell-mediated antimicrobial defence and altered expression of differentiation markers in the small intestine (Extended Data Fig. 2b, c). On exposure to bacteria or bacterial antigens, Paneth cells secrete defensins (cryptdins), which have antibacterial function during immune responses and thereby shape the composition of the microbiota<sup>11</sup>. This activity was significantly decreased in *K-ras*<sup>G12Dint</sup> mice regardless of the tumour incidence (Extended Data Fig. 2a, c). The intestinal epithelium is covered by a mucus layer, which is largely composed of mucins and provides a physical barrier, thereby limiting damage to the epithelium and enhancing gut homeostasis by delivering tolerogenic signals<sup>12</sup>. Indeed, the expression of the mucin MUC2 was significantly decreased after HFD intake (Extended Data Fig. 2c). In agreement with the duodenal gene expression data, MHC class II expression in CD11c<sup>+</sup> and CD11b<sup>+</sup> cell populations in the lamina propria (Fig. 1e), as well as in the Peyer's patches (Extended Data Fig. 2d), was significantly reduced in *K-ras*<sup>G12Dint</sup> mice regardless of diet. Taken together, these data suggest that diminished cryptdin expression by Paneth cells, owing to oncogene activation in combination with altered mucin profiles caused

<sup>1</sup>Institute of Molecular Immunology, Klinikum rechts der Isar, Technical University Munich, 81675 Munich, Germany. <sup>2</sup>Department of Molecular Biology and Genetics, Bogazici University, 34342 Bebek, Istanbul, Turkey. <sup>3</sup>Georg-Speyer-Haus, Institute for Tumor Biology and Experimental Therapy, 60596 Frankfurt am Main, Germany. <sup>4</sup>German Cancer Consortium (DKTK), 69120 Heidelberg, Germany. <sup>5</sup>German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany. <sup>6</sup>Department of Internal Medicine II, Universitätsmedizin Mannheim, Medical Faculty Mannheim, Heidelberg University, 68167 Mannheim, Germany. <sup>7</sup>Microarray and Deep-Sequencing Core Facility, University Medical Center Göttingen, 37077 Göttingen, Germany. <sup>8</sup>Institute for Mathematical Statistics, Technical University Munich, 81675 Munich, Germany. <sup>9</sup>Department of Gastrointestinal Microbiology, German Institute of Human Nutrition Potsdam-Rehbruecke, 14558 Nuthetal, Germany. <sup>10</sup>Center for Bioinformatics and Computational Biology, Delaware Biotechnology Institute, University of Delaware, Newark, Delaware 19711, USA. <sup>11</sup>Institute of Pathology, Ludwig Maximilians University, 80337 Munich, Germany. †Present address: Frutarom Savory Solutions, 49451 Holdorf, Germany.

\*These authors contributed equally to this work.



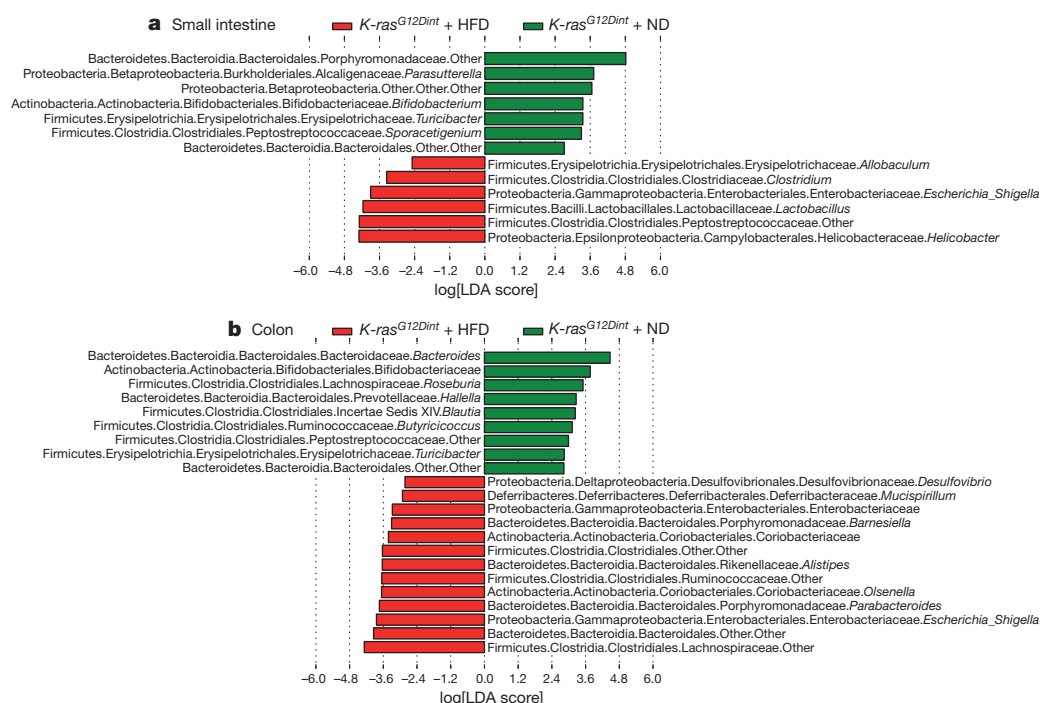
**Figure 1 | An HFD accelerates cancer progression.** **a**, The experimental scheme shows when and how long the special diet (HFD) was administered before final analysis. ND, normal diet. **b**, Histological scores show tumour incidence in the small intestine with the HFD regimen. Each point represents one animal, and lines indicate the mean (ND, *LSL-K-ras<sup>G12D/+</sup>* controls  $n = 5$ , *K-ras<sup>G12D/+</sup>* mice,  $n = 6$ ; HFD, *LSL-K-ras<sup>G12D/+</sup>* controls  $n = 13$ , *K-ras<sup>G12D/+</sup>* mice  $n = 16$ ). The scores were assessed using a two-sided Fisher test, and adjustments on pairwise  $t$ -tests were made using the single-step method.  $^*P \leq 0.05$ . **c**, Histology of the small intestine from *K-ras<sup>G12D/+</sup>* mice during murine serrated adenoma (mSA) development and comparison with the human serrated route. HGD, high-grade dysplasia; LGD, low-grade dysplasia;

mSH, murine serrated hyperplasia; TSA-HGIEN, traditional serrated adenoma with high-grade intraepithelial neoplasia; TSA-LGIEN, traditional serrated adenoma with low-grade intraepithelial neoplasia. **d**, Gene expression profiles of representative top candidate genes analysed by microarray analysis in duodenal samples from littermate *LSL-K-ras<sup>G12D/+</sup>* controls and *K-ras<sup>G12D/+</sup>* mutants on a ND or HFD regimen ( $n = 3$  per group).  $P$  values were obtained from the moderated  $t$ -statistic and corrected for multiple testing with the Benjamini-Hochberg method. **e**, Representative flow cytometric analysis showing MHC class II expression in CD11c<sup>+</sup> and CD11b<sup>+</sup> cell populations from the lamina propria (LP) in littermates ( $n = 2$  per group). Percentages do not add to 100% because of rounding errors.

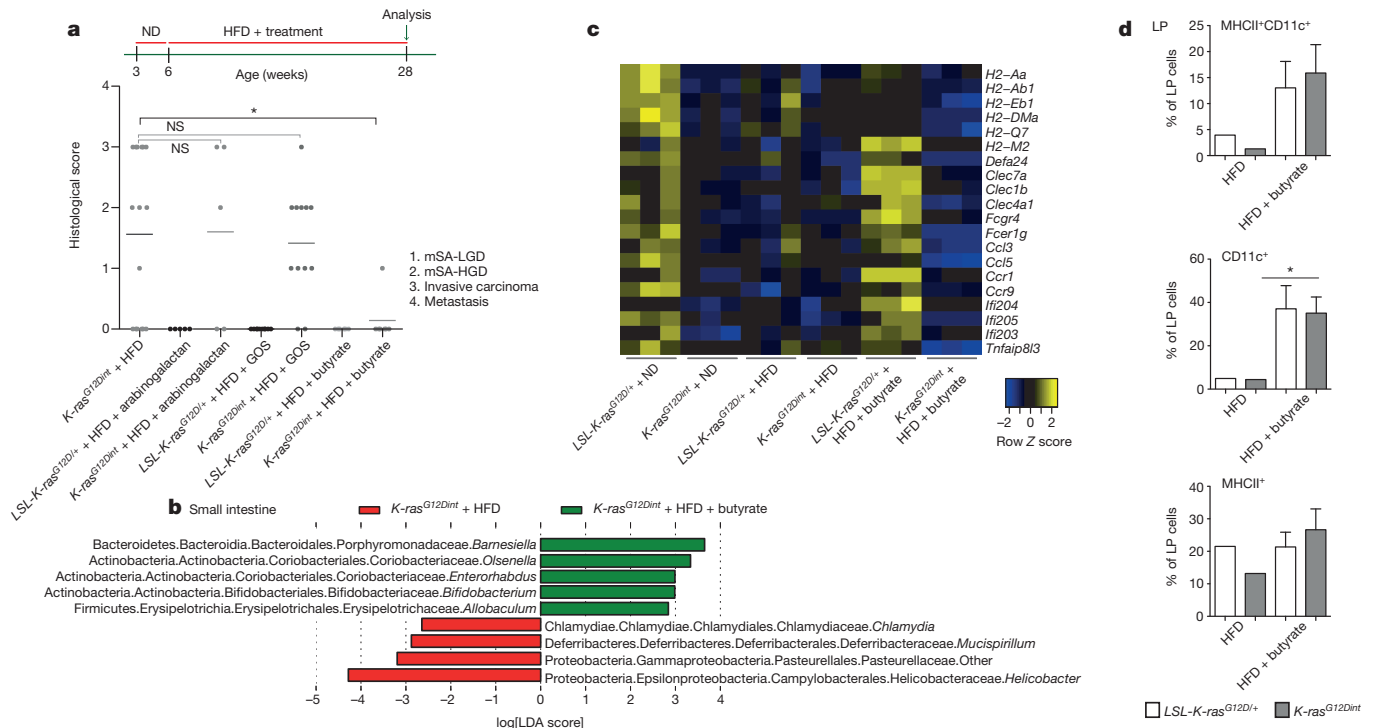
by an HFD, could render *K-ras<sup>G12D/+</sup>* mice susceptible to dampened immunity.

An HFD can lead to marked changes in the secretory, absorptive and immune function of the gut by regulating microbial communities<sup>13</sup>. To

define whether HFD-induced alterations in the microbiota were associated with the increased tumour incidence, amplicons generated from small intestinal and colonic faecal DNA were subjected to 16S ribosomal RNA gene sequencing. The HFD altered the community diversity in



**Figure 2 | Diet-induced tumour progression is associated with an altered microbial community.** Linear discriminant analysis (LDA) effect size (LEfSe) results show bacteria that were significantly different in abundance between *K-ras<sup>G12D/+</sup>* mice on a ND and an HFD regimen, and they indicate the effect size of each differentially abundant bacterial taxon in the small intestine (ND,  $n = 3$ ; HFD,  $n = 8$ ) (**a**) and the colon (ND,  $n = 3$ ; HFD,  $n = 7$ ) (**b**).



**Figure 3 | Butyrate supplementation, but not prebiotics, confers protection against HFD-induced tumorigenesis.** **a**, The experimental scheme indicates additional nutritional supplementation during the HFD and the time point of data analysis. Histological scores for the small intestine in *K-ras*<sup>G12D/+</sup> mice on an HFD (*n* = 16), after arabinogalactan supplementation (*LSL-K-ras*<sup>G12D/+</sup> controls, *n* = 5; *K-ras*<sup>G12D/+</sup> mice, *n* = 5), after galacto-oligosaccharide (GOS) supplementation (*LSL-K-ras*<sup>G12D/+</sup> controls, *n* = 10; *K-ras*<sup>G12D/+</sup> mice, *n* = 12) and after butyrate supplementation (*LSL-K-ras*<sup>G12D/+</sup> controls, *n* = 6; *K-ras*<sup>G12D/+</sup> mice, *n* = 7). Each point represents one animal, and the lines indicate the mean. The scores were assessed using a one-way analysis of variance (ANOVA), and adjustments on pairwise *t*-tests were made using the single-step method. \*, *P* ≤ 0.05; NS, not significant. **b**, LefSe results show the bacteria that were significantly different in abundance between *K-ras*<sup>G12D/+</sup> mice on an HFD that were treated (*n* = 5) or not treated (*n* = 8) with butyrate,

the intestine compared with the normal diet. The abundance of Helicobacteraceae, Lactobacillaceae, Enterobacteriaceae, Clostridiaceae and Peptostreptococcaceae was significantly higher in the small intestine of *K-ras*<sup>G12D/+</sup> mice fed an HFD than in those fed a normal diet, whereas the abundance of Bifidobacteriaceae, Porphyromonadaceae and Alcaligenaceae was significantly lower (Fig. 2a). Moreover, although no tumours were detected in the colon, the abundance of Enterobacteriaceae, Desulfovibrionaceae, Porphyromonadaceae, Rikenellaceae, Ruminococcaceae, Lachnospiraceae, Coriobacteriaceae and Deferribacteraceae was significantly higher in the colonic tissue of *K-ras*<sup>G12D/+</sup> mice fed an HFD than in those fed a normal diet, whereas the abundance of Bifidobacteriaceae, Peptostreptococcaceae, *Roseburia* and *Butyrivibrio* was significantly lower (Fig. 2b). Overall, the HFD initiated a major structural change in the gut microbiota in tumour-bearing *K-ras*<sup>G12D/+</sup> mice (Fig. 2 and Extended Data Fig. 3a–c).

Bacteria are sensed through pattern recognition receptors, Toll-like receptors and the downstream adaptor protein MYD88 (ref. 14). Interestingly, systemic ablation of *Myd88* conferred complete protection against tumour progression, suggesting a causal role for intestinal microbiota in the tumorigenic process (Extended Data Fig. 4a). To distinguish which *Myd88*-deficient cell type conferred protection against tumorigenesis, *K-ras*<sup>G12D/+</sup> mice were either crossed to *Myd88*<sup>fl/fl</sup> animals or transplanted with *Myd88*-deficient bone marrow to restrict *Myd88* deletion to intestinal epithelial cells (IECs) or haematopoietic cells, respectively. *K-ras*<sup>G12D/+</sup> mice continued to develop invasive cancer when *Myd88* deletion was restricted to the IECs (Extended Data Fig. 4a). Although

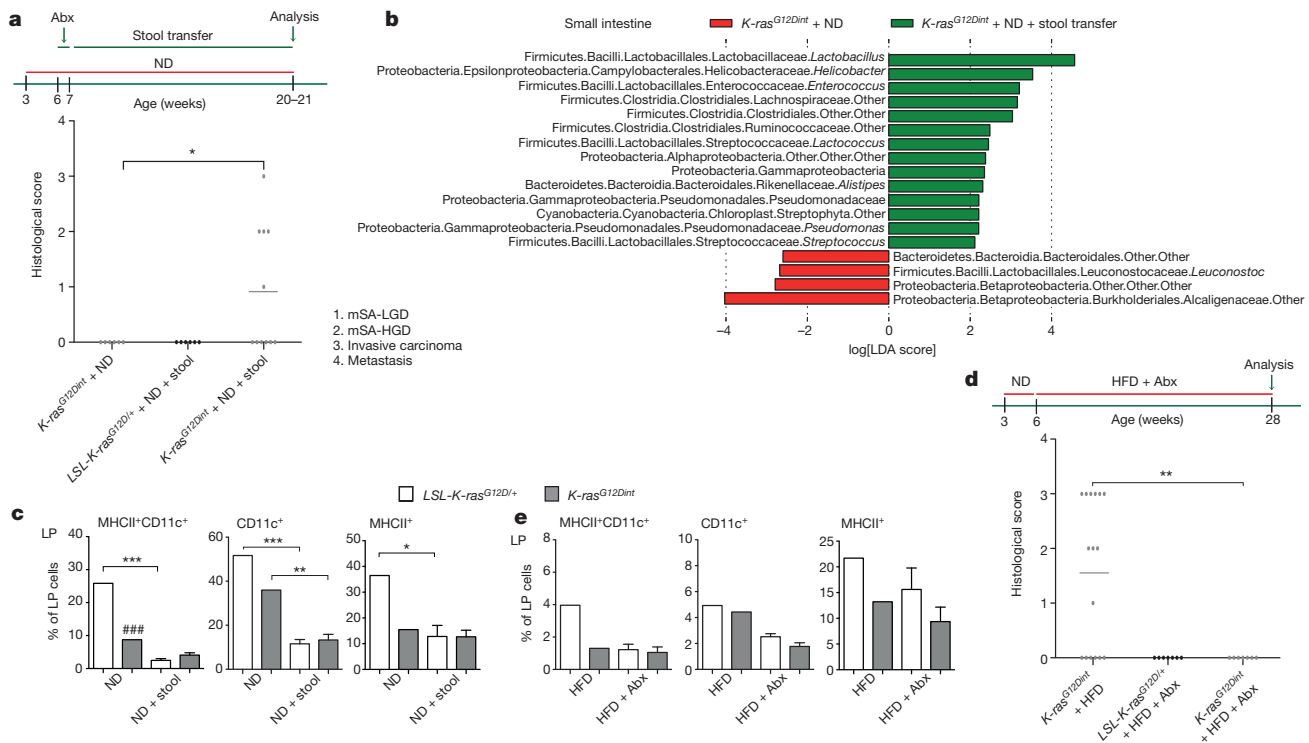
and they indicate the effect size of each differentially abundant bacterial taxon in the small intestine. **c**, The gene expression profiles of selected top candidate genes involved in antigen recognition, the immune response and immune cell recruitment in duodenal samples from *LSL-K-ras*<sup>G12D/+</sup> controls and littermate *K-ras*<sup>G12D/+</sup> animals (*n* = 3 per group) that were on the ND or HFD regimen and were treated or not treated with butyrate. *P* values were obtained from the moderated *t*-statistic and corrected for multiple testing with the Benjamini–Hochberg method. **d**, Flow cytometric analysis of LP cells from *LSL-K-ras*<sup>G12D/+</sup> control and *K-ras*<sup>G12D/+</sup> littermate animals on an HFD treated with butyrate (*LSL-K-ras*<sup>G12D/+</sup> controls, *n* = 4; *K-ras*<sup>G12D/+</sup> mice, *n* = 7) or not treated with butyrate (*LSL-K-ras*<sup>G12D/+</sup> controls, *n* = 2; *K-ras*<sup>G12D/+</sup> mice, *n* = 2). *P* values were determined by one-way ANOVA and adjusted for the number of comparisons by using the Bonferroni method. Error bars indicate s.e.m. \*, *P* ≤ 0.05.

the adoptive transfer of *Myd88*-deficient bone marrow slightly reduced tumour incidence, it did not prevent tumour progression, as the number of invasive cancers was comparable to that in mice that had received wild-type bone marrow cells (Extended Data Fig. 4a).

Defective innate immunity has been suggested to shape a distinct intestinal microbiota in mice<sup>15–17</sup>. To test whether the discrepancy in tumour incidence between germline and tissue-specific *Myd88* deficiency was due to a change in the gut microbiota, we surveyed microbial community diversity. Systemic deletion of *Myd88* led to an increase in Peptostreptococcaceae, Deferribacteraceae, and butyrate-producing Ruminococcaceae (Extended Data Fig. 4b). Furthermore, Enterococcaceae, the abundance of which was consistently increased after tissue-specific *Myd88* deletion, was not detectable when *Myd88* was deleted from the germ line (Extended Data Fig. 5a–c). Remarkably, oncogene- and diet-uncoupled antigen presentation in dendritic cells of the lamina propria was significantly attenuated after systemic *Myd88* deletion (Extended Data Fig. 4c). A possible explanation for the differential susceptibility to cancer is that in systemic deletion, the deletion of *Myd88* in IECs and haematopoietic cells has an additive protective effect. Alternatively, systemic *Myd88* deficiency might have prevented tumour progression by mechanisms that are not only associated with community changes or mucosal immunity but rather by altering IEC differentiation during embryogenesis.

The ingestion of dietary fibre promotes short-chain fatty acid (SCFA) formation and has beneficial effects on health by selectively affecting a restricted number of bacterial genera (*Bifidobacterium* and *Lactobacillus*)





**Figure 4** | Disease-associated bacteria can be transmitted to healthy  $K\text{-ras}^{G12Dint}$  animals, and antibiotic treatment abolishes tumorigenesis.

**a**, The experimental scheme indicates the time of stool transfer from HFD-fed donors that had been on the diet regimen for 24 weeks to healthy 7-week-old  $K\text{-ras}^{G12Dint}$  recipients on a ND after 1 week of antibiotic (Abx) treatment. The histological score of the small intestine suggested that a disease-associated microbiota had been delivered, as well as increased tumour incidence in the recipient  $K\text{-ras}^{G12Dint}$  mice ( $LSL\text{-K-ras}^{G12D/+}$  controls,  $n = 6$ ;  $K\text{-ras}^{G12Dint}$  mice,  $n = 11$ ). Each point represents one animal, and the lines indicate means. The scores between  $K\text{-ras}^{G12Dint}$  mice treated or not treated with stool samples were assessed using a two-sided, two-sample Welch's  $t$ -test. \*,  $P \leq 0.05$ . **b**, LEfSe results show bacteria that were significantly different in abundance between ND-fed  $K\text{-ras}^{G12Dint}$  mice that underwent stool transfer ( $n = 6$ ) or did not undergo stool transfer (ND,  $LSL\text{-K-ras}^{G12D/+}$  controls  $n = 2$ ,  $K\text{-ras}^{G12Dint}$  mice  $n = 2$ ; ND + stool transfer,  $LSL\text{-K-ras}^{G12D/+}$  controls  $n = 7$ ,  $K\text{-ras}^{G12Dint}$  mice  $n = 10$ ).  $P$  values were determined by one-way ANOVA and adjusted for the

number of comparisons by using the Bonferroni method. Error bars indicate s.e.m. \*,  $P \leq 0.05$ ; \*\*,  $P \leq 0.01$ ; \*\*\*,  $P \leq 0.001$ ; ###,  $P \leq 0.001$  compared with littermate controls. **d**, The experimental scheme indicating the application of the antibiotic regimen during the course of the HFD and the time point for data analysis. The histological scores of the small intestine samples showed complete protection against tumour formation in the HFD-fed  $K\text{-ras}^{G12Dint}$  mice treated with antibiotics (HFD + Abx,  $LSL\text{-K-ras}^{G12D/+}$  controls,  $n = 7$ ;  $K\text{-ras}^{G12Dint}$  mice,  $n = 7$ ). Each point represents one animal, and the lines indicate means. The scores were assessed using a two-sided Fisher test, and adjustments on pairwise  $t$ -tests were made using the single-step method. \*\*,  $P \leq 0.01$ . **e**, Flow cytometric analysis of LP cells showing the recruitment of, and surface antigen presentation on, CD11c<sup>+</sup> dendritic cells in disease-free  $K\text{-ras}^{G12Dint}$  mice after treatment with antibiotics (HFD + Abx,  $LSL\text{-K-ras}^{G12D/+}$  controls  $n = 5$ ,  $K\text{-ras}^{G12Dint}$  mice  $n = 7$ ; HFD,  $LSL\text{-K-ras}^{G12D/+}$  controls  $n = 2$ ,  $K\text{-ras}^{G12Dint}$  mice  $n = 2$ ).  $P$  values were determined by one-way ANOVA and adjusted for the number of comparisons by using the Bonferroni method. Error bars indicate s.e.m. The differences are not significant.

in the gut<sup>18</sup>. HFD-induced tumour progression was correlated with a significant reduction in acetate, propionate and butyrate concentrations in faecal samples (Extended Data Fig. 6a, b). To ascertain whether diet-induced tumorigenesis could be halted through further changes in the microbiota, prebiotics were administered to mice. Interestingly, arabinogalactan supplementation did not provide any protective effect (Fig. 3a). However, different prebiotics are available, with the key differentiating factor being the length of the chemical chain, which determines where in the gastrointestinal tract the prebiotic exerts its effect<sup>19</sup>. Supplementation of the diet with galacto-oligosaccharide (GOS) did not affect tumour incidence (Fig. 3a) but slightly increased the number of tumours per mouse. CD11c<sup>+</sup> dendritic cell numbers and MHC class II presentation on CD11c<sup>+</sup> dendritic cells in the lamina propria and mesenteric lymph nodes (MLNs) (Extended Data Fig. 6c), as well as expression of the genes involved in the immune response (Extended Data Fig. 7a), were equally reduced in HFD-fed  $K\text{-ras}^{G12Dint}$  mice that had been treated with GOS and those that had not (compare with Extended Data Fig. 2). The lack of a protective effect could be due to the prebiotics having little or no effect on SCFA production throughout the gut (Extended Data Fig. 7b). In addition, SCFA levels were significantly lower in the small intestine than in colonic and faecal samples (Extended Data Fig. 6b).

Therefore, we next sought to determine whether diet-induced tumour progression could be prevented when mice were orally treated with butyrate. SCFAs, the end products of colonic bacterial fermentation, have several beneficial effects on the host. They serve as an energy source, modulate intestinal motility, are a defence barrier and have been suggested to have immunoregulatory functions<sup>20,21</sup>. Although butyrate supplementation led to only a minor increase in the faecal butyrate concentration (Extended Data Fig. 8a), it markedly reduced the tumour incidence (Fig. 3a). This decrease was associated with a significant increase in the abundance of Bifidobacteriaceae and Porphyromonadaceae and a sharp decrease in Helicobacteraceae in the small intestine (Fig. 3b). Interestingly, branching, serration and proliferation were partially blocked in  $K\text{-ras}^{G12Dint}$  mice on treatment with butyrate (Extended Data Fig. 8b). The HFD-induced decrease in the expression of *Muc2* (Extended Data Fig. 8c) and genes involved in antigen recognition and the immune response (Fig. 3c) was partially restored towards the level of the normal diet group. Moreover, compromised dendritic cell recruitment was nearly normalized (Fig. 3d). Although most studies have focused on colonic bacteria, the small intestine has a crucial role in carbohydrate and fat uptake<sup>22</sup>. However, the concentration of SCFAs in the small intestine remained lower than that in colonic or faecal samples (Extended Data Fig. 6b). Butyrate seemed to be a potent SCFA, with systemic effects on



metabolic parameters during diet-induced obesity (Extended Data Fig. 8d), although histone H3 and H4 acetylation remained unaffected (Extended Data Fig. 8e). These results indicate that butyrate exerts its protective effect on tumorigenesis at least in part through changes in bacterial composition and through regulating K-Ras signalling.

To confirm the causal relationship between diet-induced dysbiosis and intestinal cancer, mice on the normal diet were treated for 1 week with an antibiotic mixture and then colonized with fresh faecal samples from HFD-fed *K-ras*<sup>G12Dint</sup> donors. Remarkably, disease was transmitted to otherwise healthy *K-ras*<sup>G12Dint</sup> mice on a normal diet but not to *LSL-K-ras*<sup>G12D/+</sup> controls (Fig. 4a and Extended Data Fig. 9a). Comparison of the microbial community composition between *K-ras*<sup>G12Dint</sup> mice that did and did not undergo stool transfer showed a higher abundance of Lactobacillaceae, Helicobacteraceae and Clostridiales in stool recipients, reflecting the transfer of the HFD-shaped microbiota to the normal diet phenotype (Fig. 4b and Extended Data Fig. 9b). MUC2 expression was significantly lower in mice that received the stool than in those that did not, reminiscent of the HFD group (Extended Data Fig. 10a). The recruitment of dendritic cells and MHC class II presentation by dendritic cells in the lamina propria and Peyer's patches were also compromised in mice that received the stool (Fig. 4c and Extended Data Fig. 9c), and glucose clearance during a glucose tolerance test was similar to that of the normal diet group (compare Extended Data Fig. 9d with Fig. 1d). These findings provide evidence that a diet-shaped microbiota synergizes with oncogenic *K-ras* during tumorigenesis in the intestine, independently of obesity.

To further support the evidence that a distinct shift in bacterial community has a causative role in tumour progression, the mice were treated with antibiotics. In support of the stool transfer experiments and the existence of a disease-associated microbiota, antibiotic supplementation abolished tumour formation in *K-ras*<sup>G12Dint</sup> mice (Fig. 4d). The HFD-mediated downregulation of genes involved in the immune response, the mucin profile and the differentiation markers in the duodenum were less pronounced after antibiotic treatment (Extended Data Fig. 7a). MHC class II presentation by dendritic cells was still compromised in the lamina propria after antibiotic treatment (Fig. 4e) but not in the MLNs (Extended Data Fig. 10b). Taken together, these results suggest a critical role for diet-shaped dysbiotic bacteria in aggravating oncogene-driven intestinal carcinogenesis (Extended Data Fig. 10c).

The perturbation of immunoregulatory functions by a dysbiotic microbiota is becoming increasingly recognized as a hallmark of immune-mediated diseases<sup>3</sup>. However, diet-associated cancer development may be based on marked shifts in bacterial communities rather than on the development of obesity and metabolic disorder. Thus, personalized dietary interventions might allow an individual's microbiota to be modulated to promote health, especially in those who are at a high risk because of genetic susceptibility and a high fat intake.

## METHODS SUMMARY

*K-ras*<sup>G12Dint</sup> mice on a C57BL/6J;129 background have been described previously<sup>8</sup>. The mice were fed either with a total pathogen-free diet (Altromin, catalogue no. 1314; percentage of total calories, 27% protein, 60% carbohydrate and 13% fat) (Supplementary Table 1) or a  $\gamma$ -irradiated HFD (Research Diets, catalogue no. D12492; percentage of total calories, 20% protein, 20% carbohydrate and 60% fat) (Supplementary Table 2). Sections from the small intestine and colon were evaluated and scored by pathologists in a blinded manner. For immunohistochemistry, paraffin-embedded sections were stained according to standard procedures using a monoclonal rat anti-5-bromodeoxyuridine (BrdU) antibody (1:200; AbD Serotec, catalogue no. MCA 2060) and counterstained with haematoxylin. BrdU (100 mg per kg body weight; Sigma, catalogue no. B9285) was intraperitoneally injected 2 h before the mice were killed.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 9 December 2012; accepted 22 April 2014.

Published online 31 August 2014.

1. Giovannucci, E. & Michaud, D. The role of obesity and related metabolic disturbances in cancers of the colon, prostate, and pancreas. *Gastroenterology* **132**, 2208–2225 (2007).

- Wu, G. D. *et al.* Linking long-term dietary patterns with gut microbial enterotypes. *Science* **334**, 105–108 (2011).
- Schwabe, R. F. & Jobin, C. The microbiome and cancer. *Nature Rev. Cancer* **13**, 800–812 (2013).
- Turnbaugh, P. J., Backhed, F., Fulton, L. & Gordon, J. I. Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. *Cell Host Microbe* **3**, 213–223 (2008).
- Zimmer, P., Alberti, K. G. & Shaw, J. Global and societal implications of the diabetes epidemic. *Nature* **414**, 782–787 (2001).
- Ley, R. E., Turnbaugh, P. J., Klein, S. & Gordon, J. I. Microbial ecology: human gut microbes associated with obesity. *Nature* **444**, 1022–1023 (2006).
- Turnbaugh, P. J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027–1031 (2006).
- Bennecke, M. *et al.* *Ink4a/Arf* and oncogene-induced senescence prevent tumor progression during alternative colorectal tumorigenesis. *Cancer Cell* **18**, 135–146 (2010).
- Arkan, M. C. *et al.* IKK- $\beta$  links inflammation to obesity-induced insulin resistance. *Nature Med.* **11**, 191–198 (2005).
- Khasawneh, J. *et al.* Inflammation and mitochondrial fatty acid  $\beta$ -oxidation link obesity to early tumor promotion. *Proc. Natl Acad. Sci. USA* **106**, 3354–3359 (2009).
- Clevers, H. C. & Bevins, C. L. Paneth cells: maestros of the small intestinal crypts. *Annu. Rev. Physiol.* **75**, 289–311 (2013).
- Shan, M. *et al.* Mucus enhances gut homeostasis and oral tolerance by delivering immunoregulatory signals. *Science* **342**, 447–453 (2013).
- Maslowski, K. M. & Mackay, C. R. Diet, gut microbiota and immune responses. *Nature Immunol.* **12**, 5–9 (2011).
- Kawai, T. & Akira, S. The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. *Nature Immunol.* **11**, 373–384 (2010).
- Slack, E. *et al.* Innate and adaptive immunity cooperate flexibly to maintain host-microbiota mutualism. *Science* **325**, 617–620 (2009).
- Larsson, E. *et al.* Analysis of gut microbial regulation of host gene expression along the length of the gut and regulation of gut microbial ecology through MyD88. *Gut* **61**, 1124–1131 (2012).
- Ubeda, C. *et al.* Familial transmission rather than defective innate immunity shapes the distinct intestinal microbiota of TLR-deficient mice. *J. Exp. Med.* **209**, 1445–1456 (2012).
- Redgwell, R. J. & Fischer, M. Dietary fiber as a versatile food component: an industrial perspective. *Mol. Nutr. Food Res.* **49**, 521–535 (2005).
- Macfarlane, G. T., Steed, H. & Macfarlane, S. Bacterial metabolism and health-related effects of galacto-oligosaccharides and other prebiotics. *J. Appl. Microbiol.* **104**, 305–344 (2008).
- Furusawa, Y. *et al.* Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells. *Nature* **504**, 446–450 (2013).
- Brestoff, J. R. & Artis, D. Commensal bacteria at the interface of host metabolism and the immune system. *Nature Immunol.* **14**, 676–684 (2013).
- Zoetendal, E. G. *et al.* The human small intestinal microbiota is driven by rapid uptake and conversion of simple carbohydrates. *ISME J.* **6**, 1415–1426 (2012).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank K. Burmeister and J. Khasawneh for technical assistance and H. Wagner for generously providing *Myd88*<sup>-/-</sup> mice. We are thankful to K. Offe and the PhD program 'Medical Life Sciences and Technology' for providing a fellowship to J.H. for one year. This work was supported in part by the LOEWE Center for Cell and Gene Therapy Frankfurt (CGT, III L 4-518/17.004) and institutional funds from the Georg-Speyer-Haus, as well as grants from the Deutsche Forschungsgemeinschaft (DFG) (Gr1916/5-1), the Deutsche Krebshilfe (108872) and the ERC (ROSCAN-281967) to F.R.G. Computational infrastructure made available to S.W.P. by the University of Delaware Center for Bioinformatics and Computational Biology Core Facility and the Delaware Biotechnology Institute was supported by grants from the US National Institutes of Health National Institute of General Medical Sciences (8 P20 GM103446-12) and the US National Science Foundation EPSCoR (EPS-081425). This work was supported by grants from the Deutsche Krebshilfe (107977) and the DFG (AR710/2-1) to M.C.A.

**Author Contributions** M.D.S., Ç.A., J.H. and M.C.A. performed experimental work and managed data analyses. F.K.R., S.S., B.A., P.K.Z., J.V., W.R. and F.R.G. assisted with sample collection, bone marrow transplantation and flow cytometric analyses. C.P. and G.S.-R. carried out sample preparation and microarray analyses. A.B. conducted statistical analyses. C.A. and M.B. performed gas chromatography. L.B. and T.K. provided human samples and evaluated all histological sections. S.C.P. and S.W.P. conducted 16S rRNA gene sequencing computational analyses. M.C.A. designed the study and prepared the manuscript.

**Author Information** Microarray data were generated in accordance with the MIAME guidelines and have been deposited in the Gene Expression Omnibus (GEO) database under accession number GSE56257. Sequence data have been deposited in the NCBI Sequence Read Archive (SRA) database under BioProject PRJNA242565 (SRA project accession number, SRP040736; sample accession numbers, SRS584259–SRS584323, SRS584325). Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.C.A. ([canan.arkan@lrz.tum.de](mailto:canan.arkan@lrz.tum.de)).

## METHODS

**Mice.** Littermate mice were co-housed randomly regardless of their genotype during the various diet and additional nutritional supplementation regimens. However, during stool transfer experiments, the mice were separated and co-housed according to their genotypes to avoid cross contamination. During treatment with antibiotics, the drinking water was supplemented with a mixture of 0.2 g l<sup>-1</sup> ampicillin, 0.1 g l<sup>-1</sup> vancomycin, 0.2 g l<sup>-1</sup> neomycin and 0.2 g l<sup>-1</sup> metronidazole. During transfer experiments, following antibiotic treatment for 1 week, 7-week-old mice were colonized with fresh stool pellets (9 × 10<sup>6</sup> bacteria) from mice that had been fed the HFD for 24 weeks. GOS (Bi2muno) (5.5 g) was provided in 150 ml drinking water, which was refreshed every 2 days. Sodium butyrate (Aldrich, catalogue no. 303410) (50 µg per g body weight in 100 µl water) and arabinogalactan (Fluka, catalogue no. A-09788) (0.01 per g body weight in 100 µl water) were provided orally three times a week. *K-ras*<sup>G12Dint</sup> mice were further crossed to *Myd88*<sup>fl/fl</sup> and *Myd88*<sup>-/-</sup> mice (Jackson Laboratory). During adoptive transfer experiments, 6-week-old *K-ras*<sup>G12Dint</sup> mice were irradiated (9 Gy), and 2 × 10<sup>6</sup> bone marrow cells from *Myd88*<sup>-/-</sup> or wild-type mice were transferred by tail vein injection to recipients. The procedures were approved by the Regierung von Oberbayern.

**Human samples.** Human samples were analysed after irreversible anonymization of the patients' personal data. Histological sections were microscopically evaluated by pathologists after haematoxylin and eosin staining. All procedures were performed according to the recommendation of the ethics committee of Ludwig Maximilians University.

**Glucose tolerance test (GTT).** Following 9 h of fasting, the basal glucose level was detected using a glucometer (Bayer, catalogue no. 3822850). Mice were injected with 1.5 g glucose per kg body weight (40% glucose solution, Eifelango), and blood glucose levels were recorded for up to 2 h. Plasma samples were collected simultaneously, and insulin secretion was measured with the Ultra Sensitive Rat Insulin ELISA Kit (Crystal Chem, catalogue no. 90060) using a mouse standard (Crystal Chem, catalogue no. 90070). The tests were assessed in a blinded manner.

**RNA analysis.** Total RNA was extracted from the small and large intestine using the RNeasy Mini Kit (QIAGEN). cDNA was synthesized using SuperScript II Reverse Transcriptase (Invitrogen). Real-time PCR analysis using Power SYBR Green PCR Master Mix (Applied Biosystems) was carried out on a StepOnePlus Real-Time PCR System (Applied Biosystems). The primer sequences used are as follows: F4/80, Fw 5'-CTTTGGCTATGGGCTTCCAGTC-3', Rev 5'-GCAAGGAGGA CAGAGTTTATCGTG-3'; IL-1β, Fw 5'-GTGGCTGTGGAGAAGCTGTG-3', Rev 5'-GAAGTCCACGGGAAAGACAC-3'; IL-6, Fw 5'-GTATGAACAACG ATGATGCACCTTG-3', Rev 5'-ATGGTACTCCAGAAGACCAAGAGGA-3'; TNF-α, Fw 5'-ACTCCAGGCGGTGCCTATG-3', Rev 5'-GAGCGTGGTGGCCCT-3'; sucrase-isomaltase, Fw 5'-CAACCTCGCAAACCTTTATAGT-3', Rev, 5'-TGCAGCCTCTCTACGCAA-3'; synaptophysin, Fw 5'-TTCGTGAAGG TGCTGCAGTG-3', Rev 5'-TCTCCGGTGTAGCTGCCG-3'; cryptdin, Fw 5'-C AGCCGGAAGAGAGACCAAG-3', Rev, 5'-TAGCATACCAAGATCTCTCAAC GATTC-3'; MUC1, Fw 5'-GAGCCAGGACTTCTGGTAGGCT-3', Rev 5'-GG CTTCACCAAGGCTTACGTAGT-3'; MUC2, Fw 5'-TCGCCCAAGTCGACACT CA-3', Rev 5'-GCAAATAGCCATAGTACAGTTACACAGC-3'; MUC5, Fw 5'-GATCCATCCATCCATTTCTACC-3', Rev 5'-TTGCTTATCTGACTACCAC TTGTGA-3'; REG3A, Fw 5'-GGTGAGGCTTCTTTGTGTGCC-3', Rev 5'-CT CCATTGGGTTGTTGACCC-3'; MHC class II Ab1, Fw 5'-GCGCATACGAT ATGTGACCAGAT-3', Rev 5'-GCGGTGCTCGCCCA-3'; MHC class II DMA, Fw 5'-CGTTGGTCTGTTTCATCAGCA-3', Rev 5'-ATCGACAGCTGAGATG GATGTG-3'; CCL20, Fw 5'-GGTGGCAAGCGTCTGCTC-3', Rev 5'-GCCTG GCTGCAGAGGTGA-3'; defensin, Fw 5'-TCGTTCTGTGGCCTTCC-3', Rev 5'-CCTGGCTGTTCTCAGTTTATGTC-3'; FcR, Fw 5'-TTGCTCCTTTTGG TGGAACAA-3', Rev 5'-GGACAATACCATAAAAAACAGGACA-3'; CLEC4N, Fw 5'-CCCCATGAACCAATCTT-3', Rev 5'-CAGCCCCATTTGGAAGGA-3'; CLEC7A, Fw 5'-GTGCAGTAAGCTTCTCTGGG-3', Rev 5'-TCCCGCAAT CAGATGAAG-3'; and cyclophilin, Fw 5'-ATGGTCAACCCACCGTGT-3', Rev 5'-TTCTGCTGTCTTTGGAACCTTGT-3'.

**Microarray analysis using the GeneChip Mouse Gene 1.0 ST array.** RNA isolation from three biological replicates of duodenal samples was carried out using the TRIzol (Invitrogen) method according to the manufacturer's instructions. The samples were treated with DNase I (Sigma-Aldrich). RNA quality was checked by microfluidic electrophoresis using the 2100 Bioanalyzer (Agilent Technologies), and only samples with comparable RNA integrity numbers were processed for microarray analysis.

cDNA was synthesized using 0.3 µg of total RNA. The synthesis of the double-stranded cDNA was performed using WT Target Labeling and Control Reagents (Affymetrix), and the clean-up was carried out using the GeneChip Sample Clean-up module (Affymetrix).

*In vitro* transcription was conducted using the WT Target Labeling Kit. The total amount of the reaction product was purified using the GeneChip Sample Cleanup

Module and quantified with the ND-1000 (NanoDrop, Thermo). cDNA synthesis (single strand, ss) was carried out with the WT Target Labeling Kit, and total ssDNA (5.5 µg) was enzymatically cleaved into 35–200 base-pair fragments. The degree of fragmentation and the length distribution of the ssDNA were analysed by capillary electrophoresis using the 2100 Bioanalyzer. Following fragmentation, a terminal labelling reaction (with biotin) was conducted using the WT Labeling Kit.

Biotinylated fragmented ssDNA was then hybridized onto the GeneChip Mouse Gene 1.0 ST Array (Affymetrix) according to the manufacturer's instructions. The hybridization was carried out at 45 °C in the GeneChip Hybridization Oven 640 (Affymetrix) for 16 h. Washing and staining of the arrays on the GeneChip Fluidics Station 450 (Affymetrix) were performed according to the manufacturer's recommendations. Amplification of the antibody signal and washing and staining protocols were also according to the manufacturer's instructions and were used to stain the arrays with streptavidin R-phycoerythrin (SAPE, Invitrogen). The arrays were incubated twice with SAPE solution, together with a biotinylated anti-streptavidin antibody (Vector Labs) staining step, to amplify staining. Finally, the arrays were scanned using the GeneChip Scanner 3000 7G (Affymetrix).

**Microarray data processing and statistical analysis.** AGCC Software (version 2.0, Affymetrix) was used for the extraction of intensity data, which were analysed using the affy<sup>23</sup> and Limma<sup>24</sup> packages of Bioconductor<sup>25</sup>. The analysis was carried out exactly as previously described<sup>26</sup>. The data analysis consisted of between-array normalization, probe summary, global clustering and principal components analysis (PCA) analysis, fitting of the data to a linear model and detection of differential gene expression. To ensure that the intensities had similar distributions across arrays, quantile normalization was applied to the log<sub>2</sub>-transformed intensity values as a method for between-array normalization<sup>27</sup>. Similarly to the summary of probes, a median polish procedure was chosen.

Significant changes in the expression of genes between the groups were analysed by empirical Bayes statistics, by moderating the standard errors of the estimated values<sup>28</sup>.

*P* values obtained from the moderated *t*-statistic were corrected for multiple testing with the Benjamini–Hochberg method<sup>29</sup>. *P*-value adjustments guarantee a smaller number of false positive findings by controlling the false discovery rate (fdr). For each gene, the null hypothesis suggesting that there is no differential expression between degradation levels was rejected when the fdr was lower than 0.05. Samples were assessed in a blinded manner.

**Cell isolation from lamina propria.** Intestinal tissue was cut into small pieces, shaken vigorously in 25 ml RPMI and washed three times with PBS. Tissue pieces were then incubated in 50 ml PBS containing 0.015 g dithiothreitol (DTT) and 500 µl 0.5 M EDTA with constant shaking for 20 min at 37 °C. The pieces were then collected and transferred to 10 ml RPMI containing 50 µl DNase I grade II (100 mg ml<sup>-1</sup>) and 50 µl collagenase D (100 mg ml<sup>-1</sup>) and then incubated for 25 min in a rotating incubator at 37 °C. Cells were passed through a cell strainer and centrifuged at 194g for 5 min, and the resultant pellet was resuspended in fluorescence-activated cell sorting (FACS) buffer (2% FCS in PBS).

**Cell isolation from the Peyer's patches.** The Peyer's patches were excised from the intestine. They were incubated for 20 min in a mixture of 3 ml RPMI, 15 µl collagenase D (100 mg ml<sup>-1</sup>) and 15 µl DNase I grade II (20 mg ml<sup>-1</sup>) at 37 °C. The cells were then passed through a cell strainer and incubated for 10 min at 37 °C. The reaction was stopped by the addition of 10 µl 0.5 M EDTA, and the volume was adjusted to 10 ml with RPMI. After centrifugation at 249g for 5 min, the cells were resuspended in FACS buffer.

**Cell isolation from the MLNs.** The MLNs were collected in 5 ml 2% FCS in PBS, sheared on glass slides and passed through a 100 µm filter and then a 70 µm filter. Following centrifugation at 194g for 5 min at 4 °C, the erythrocytes were lysed with 1 ml of Red Blood Cell Lysing Buffer (Sigma) for 5 min at 20 °C. The pellets were then washed in 2% FCS in PBS, centrifuged at 194g for 5 min and resuspended in FACS buffer.

**FACS.** After cell isolation from the lamina propria, Peyer's patches and MLNs, approximately 1 × 10<sup>6</sup> cells were stained with 0.5 µg ml<sup>-1</sup> ethidium monoazide (EMA) (Sigma, catalogue no. E2028) to stain dead cells for 15 min under a light source. Cells were washed and incubated with mouse BD Fc Block (BD Pharmingen, clone 2.4G2, catalogue no. 553141) for 10 min on ice and then washed and centrifuged at 775g for 5 min. Fluorescently labelled antibodies (fluorescein isothiocyanate (FITC) rat anti-mouse CD11b antibody, BD Pharmingen, clone M1/70, catalogue no. 557396; anti-mouse phycoerythrin (PE)-cyanine7 CD11c, eBioscience, clone N418, catalogue no. 25-0114-82; and allophycocyanin (APC) anti-mouse MHC class II (I-A/I-E), eBioscience, clone M5/114.15.2, catalogue no. 17-5321-81) (1:200) were added and incubated for 20 min on ice. After washing, cells were fixed by incubating in fixation buffer (eBioscience) for 30 min on ice. Following a further washing step, cells were suspended in FACS buffer, filtered using 50 µm filters (Günter Keul) and sorted using the Gallios Flow Cytometer (Beckman Coulter). Data analyses were carried out using FlowJo software (version 8.8.6).

**Sample preparation and pyrosequencing.** Stool samples were freshly collected and immediately frozen in liquid nitrogen. Genomic DNA was isolated from 200 mg frozen faecal samples using a QIAamp DNA Stool Mini Kit (QIAGEN). PCR amplification of the V1–V3 region of bacterial 16S ribosomal DNA was carried out using primers (338F, 5'-ACTCCTACGGGAGGCAGCAG-3', 806R, 5'-GGAC TACCAGGTATCTAAT-3') incorporating FLX Titanium adaptors (Roche) and a sample barcode sequence. The PCR was made based on the 'Amplicon Library Preparation Method' (Roche) using a FastStart High Fidelity PCR System (Roche), 10 mM PCR Nucleotide Mix (Roche) and 10 ng stool DNA per reaction in a 25 µl reaction volume. After purification of the product on a 1.2% agarose gel, equal concentrations of amplicons were pooled from each sample. Emulsion PCR and GS FLX amplicon sequencing were carried out according to the Roche Titanium series chemistry.

Pyrosequenced amplicon libraries were screened for quality characteristics, chimeric sequences, and PCR/pyrosequencing-induced duplication artefacts using AmpliconNoise and Perseus software (version 1.28) as previously described<sup>30</sup>. Cleaned data sets were processed using the QIIME amplicon analysis pipeline (version 1.7.0)<sup>31</sup>. Briefly, sequences were clustered into distance-based (97% similarity) operational taxonomic units (OTUs) using mothur software (version 1.30.2)<sup>32</sup>. Representatives of each OTU were aligned to the Greengenes core set alignment<sup>33</sup> using PyNAST software (version 1.1)<sup>34</sup>. Taxonomic assignments for each OTU were made using the Ribosomal Database Project (RDP) Classifier (version 2.2)<sup>35</sup>.

**LEfSe results for microbiomes.** LEfSe is an algorithm for applying 16S rRNA gene data sets to detect bacterial organisms and functional characteristics that are differentially abundant between two or more microbial environments<sup>36</sup>. It emphasizes statistical significance, biological consistency and effect relevance, allowing the identification of differentially abundant features that are also consistent with biologically meaningful categories (subclasses). LEfSe first robustly identifies features that are significantly different among biological classes. It then carries out additional tests to assess whether these differences are consistent with respect to expected biological behaviour. The non-parametric factorial Kruskal–Wallis (KW) rank-sum test is used to detect features with significant differential abundance with respect to the class of interest; biological consistency is subsequently investigated using a set of pairwise tests among subclasses using the (unpaired) Wilcoxon rank-sum test. As a last step, LEfSe uses linear discriminant analysis (LDA) to estimate the effect size of each differentially abundant feature. Samples were assessed in a blinded manner.

**SCFA measurement.** Stool samples were freshly collected and immediately frozen in liquid nitrogen. A 1:5 dilution of the samples in double distilled water was centrifuged, and the supernatant was mixed with 12 mM isobutyric acid, 1 M NaOH and 0.36 M HClO<sub>4</sub>. After lyophilization for 16 h, the remaining powder was diluted with acetone and 5 M formic acid and centrifuged, and the supernatant was used for the measurement with an HP 5890 Series II gas chromatograph (Hewlett Packard). Samples were assessed in a blinded manner.

**Histone acetylation.** Histones were extracted using an EpiQuik Total Histone Extraction Kit (Epigentek, catalogue no. OP-0006). H3 or H4 acetylation was determined using an EpiQuik Global Histone H3 (or H4) Acetylation Assay Kit (Epigentek, catalogue no. P-4008 and P-4009), according to the manufacturer's instructions.

**Statistical analysis.** Outcomes on sufficiently metric scales (cell percentage, messenger RNA levels, SCFA levels, percentage histone acetylation) were assessed with linear models (one-way ANOVA), modelling the mean value of each group, which were defined through mutation, diet and diet supplementation. Relevant comparisons of mean values between groups were made with two-sided pairwise *t*-tests; the null hypotheses stated that there was no difference in the means. *P* values were adjusted for the number of comparisons by using the Bonferroni method.

Outcomes on a discrete/ordinal scale (histological scores) were either assessed using Fisher's exact test, Welch's *t*-test or one-way ANOVA models, as above.

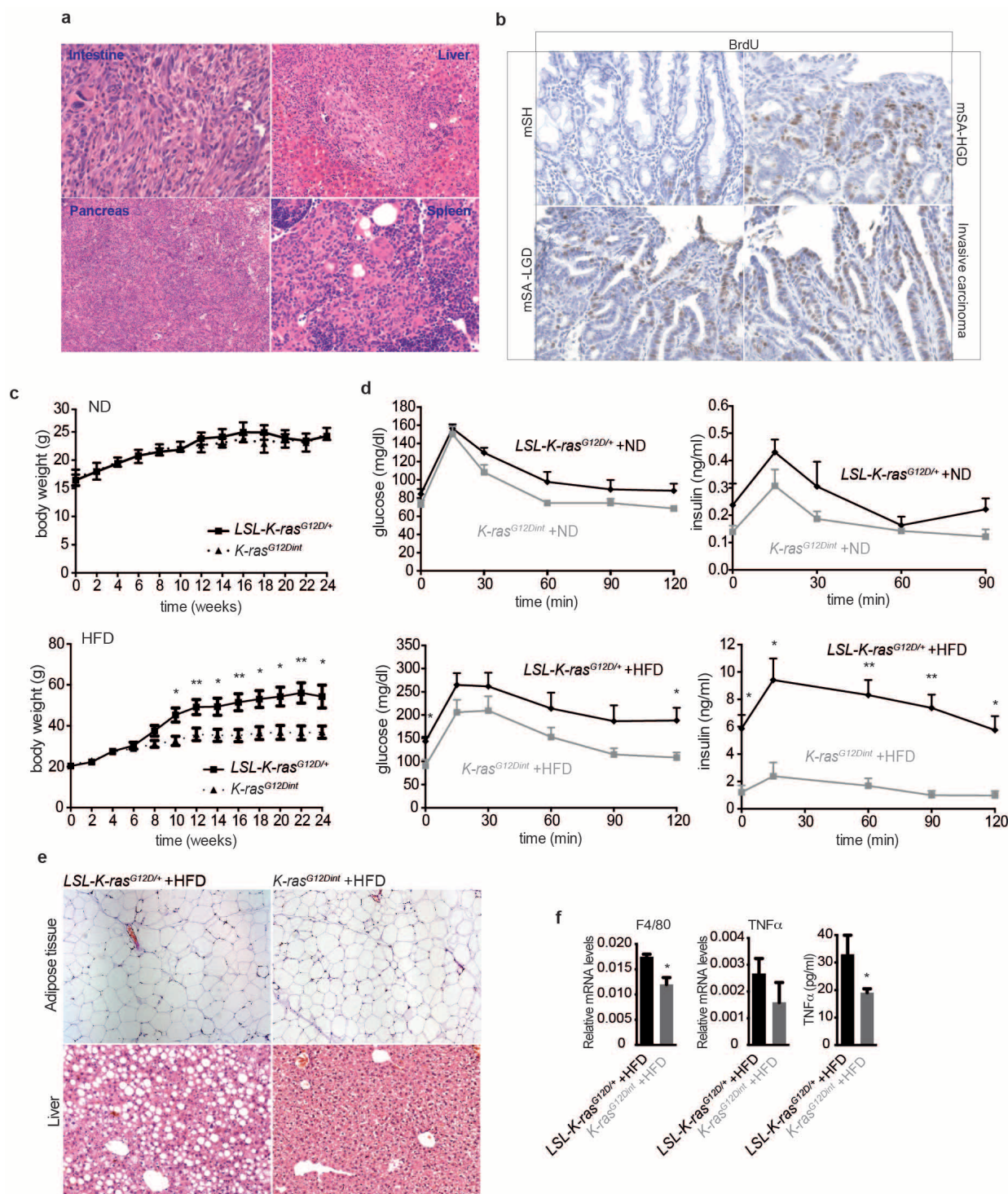
Fisher's test was chosen when one of the groups comprised only values on a single outcome level, because in such a case the assumption of equal variances is hardly violated. A two-sided Fisher's test was applied to a two-by-two contingency table classifying the histological scores as either no pathology or any pathology. The ANOVA method was chosen when both groups had observations on at least two levels of the histological score and mimicked the Cochran–Mantel–Haenszel tests for linear trends in ordinal outcomes. Adjustments on pairwise *t*-tests were made using the single-step method<sup>37</sup>. Both types of adjustments control the family-wise error rate of making one or more false discoveries.

Outcomes measured over several time points (the GTT, insulin values and the weight curve) were assessed for differences between groups (*LSL-K-ras*<sup>G12D/+</sup> control mice and *K-ras*<sup>G12D<sup>int</sup></sup> mice) with linear models. The models estimated the mean values of each group–time combination, and *t*-tests were carried out for every pair of mean values at a specific time point. *P* values were adjusted for multiple testing within each time series: that is, for 5 to 6 tests for glucose and insulin, and for 12 tests for the weight measurements. Adjustment was carried out with the single-step method as mentioned above, which is implemented in the glht (general linear hypothesis) function in the R package multcomp.

The number of mice per group used in an experiment is annotated in the corresponding figure legend as *n*. Although no prior sample size estimation was performed, we have used as many mice per group as possible. No gender differences were observed. The significance of tests is reported as NS (not significant), \* (*P* ≤ 0.05), \*\* (*P* ≤ 0.01) and \*\*\* (*P* ≤ 0.001). Statistical tests were carried out using Prism4 (GraphPad Software) and R (R Core Team, 2013).

23. Gautier, L., Cope, L., Bolstad, B. M. & Irizarry, R. A. affy—analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20**, 307–315 (2004).
24. Wettenhall, J. M. & Smyth, G. K. limmaGUI: a graphical user interface for linear modeling of microarray data. *Bioinformatics* **20**, 3705–3706 (2004).
25. Gentleman, R. C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5**, R80 (2004).
26. Meyer, S., Nolte, J., Opitz, L., Salinas-Riester, G. & Engel, W. Pluripotent embryonic stem cells and multipotent adult germline stem cells reveal similar transcriptomes including pluripotency-related genes. *Mol. Hum. Reprod.* **16**, 846–855 (2010).
27. Irizarry, R. A. *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249–264 (2003).
28. Smyth, G. K. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* <http://dx.doi.org/10.2202/1544-6115.1027> (12 February 2004).
29. Klipper-Aurbach, Y. *et al.* Mathematical formulae for the prediction of the residual β cell function during the first two years of disease in children and adolescents with insulin-dependent diabetes mellitus. *Med. Hypotheses* **45**, 486–490 (1995).
30. Quince, C., Lanzen, A., Davenport, R. J. & Turnbaugh, P. J. Removing noise from pyrosequenced amplicons. *BMC Bioinformatics* **12**, 38 (2011).
31. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7**, 335–336 (2010).
32. Schloss, P. D. *et al.* Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75**, 7537–7541 (2009).
33. DeSantis, T. Z. *et al.* Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* **72**, 5069–5072 (2006).
34. Caporaso, J. G. *et al.* PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* **26**, 266–267 (2010).
35. Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73**, 5261–5267 (2007).
36. Segata, N. *et al.* Metagenomic biomarker discovery and explanation. *Genome Biol.* **12**, R60 (2011).
37. Hothorn, T., Bretz, F. & Westfall, P. Simultaneous inference in general parametric models. *Biom. J.* **50**, 346–363 (2008).

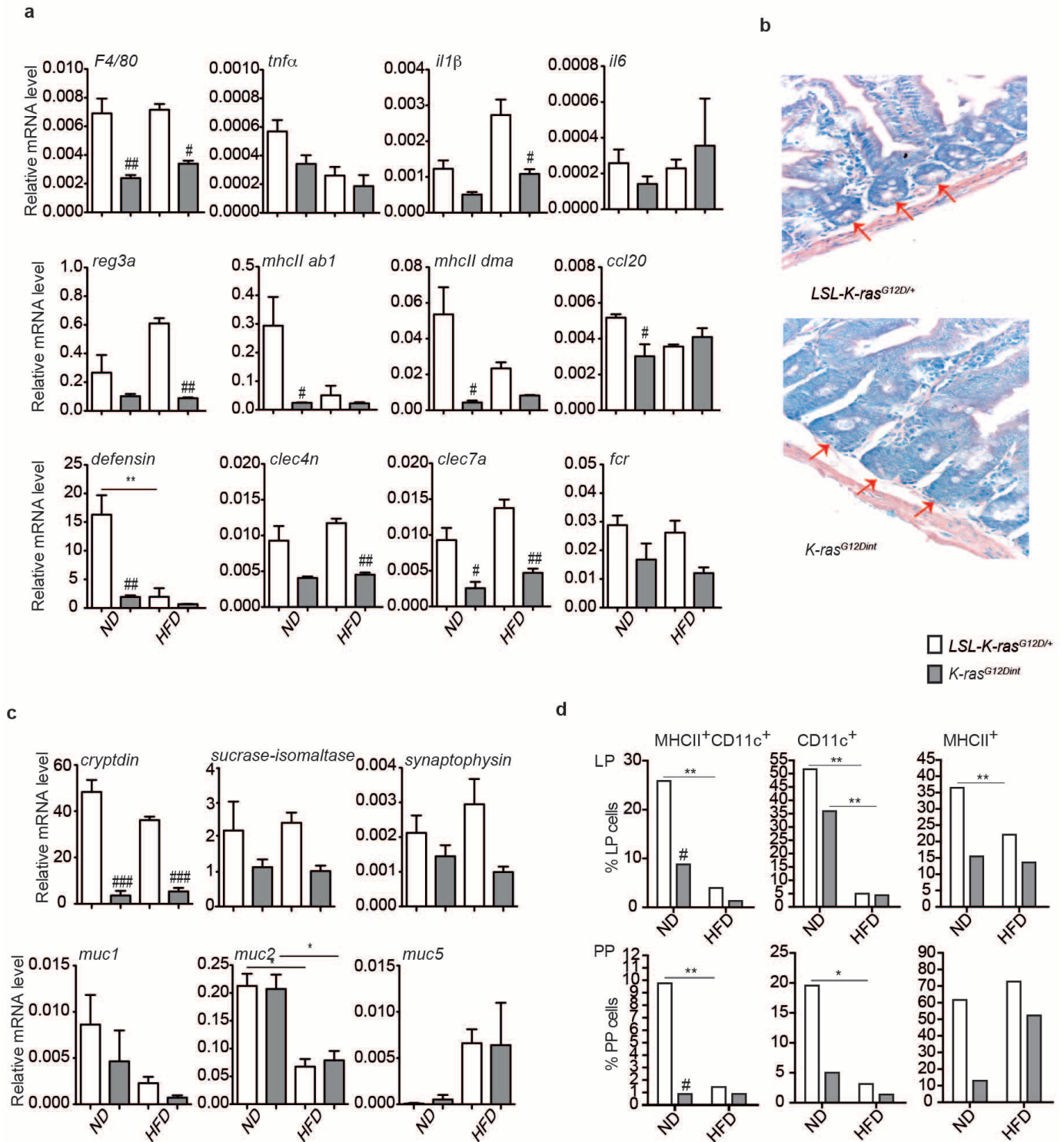




**Extended Data Figure 1 | An HFD accelerates carcinogenesis independently of obesity and insulin resistance.** **a**, Primary tumours metastasized to the liver, pancreas and spleen in *K-ras*<sup>G12Dint</sup> mice maintained on an HFD for 43 weeks. **b**, Immunohistochemistry staining of duodenal samples using an antibody specific for 5-bromodeoxyuridine (BrdU) showed increased proliferation in *K-ras*<sup>G12Dint</sup> mice. **c**, Weight curves for LSL-*K-ras*<sup>G12D/+</sup> controls ( $n = 5$ ) and *K-ras*<sup>G12Dint</sup> ( $n = 6$ ) littermate animals showed no difference in weight gain under the normal diet (ND) condition, although *K-ras*<sup>G12Dint</sup> mice ( $n = 7$ ) remained significantly leaner when fed on an HFD (LSL-*K-ras*<sup>G12D/+</sup> controls  $n = 4$ ).  $P$  values were determined by  $t$ -test and adjusted for multiple testing. The error bars indicate s.e.m. \*,  $P \leq 0.05$ ; \*\*,  $P \leq 0.01$ . **d**, In accordance with the weight curves, the response to glucose overload and insulin secretion during a glucose tolerance test (GTT) remained similar between the two groups under the ND condition; however, *K-ras*<sup>G12Dint</sup> mice remained insulin sensitive on an

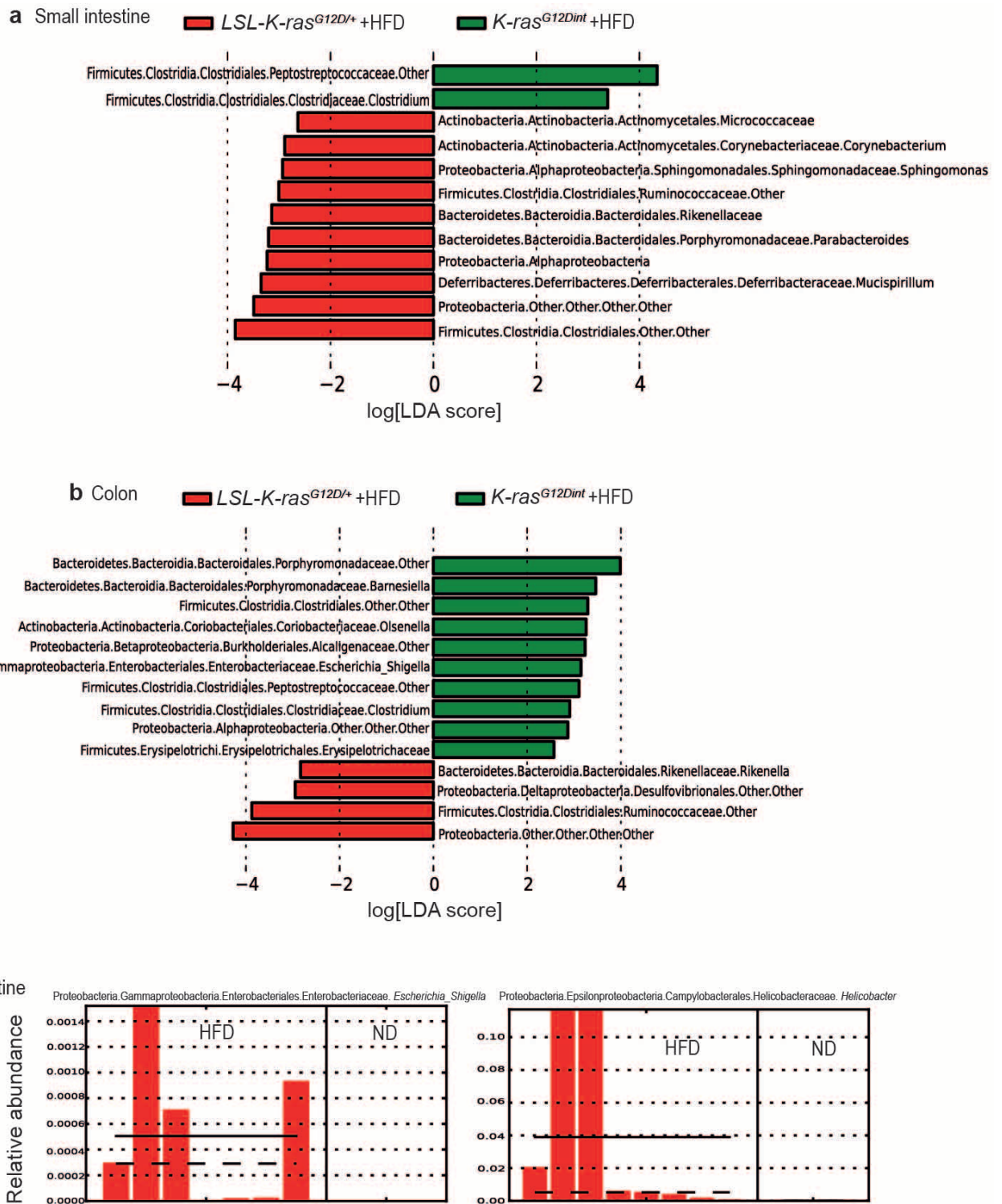
HFD (ND, LSL-*K-ras*<sup>G12D/+</sup> controls  $n = 5$ , *K-ras*<sup>G12Dint</sup> mice  $n = 3$ ; HFD, LSL-*K-ras*<sup>G12D/+</sup> controls  $n = 8$ , *K-ras*<sup>G12Dint</sup> mice  $n = 5$ ).  $P$  values were determined by  $t$ -test and adjusted for multiple testing. The error bars indicate s.e.m. \*,  $P \leq 0.05$ ; \*\*,  $P \leq 0.01$ . The results are representative of two to three independent experiments. **e**, Together with resistance to diet-induced obesity, *K-ras*<sup>G12Dint</sup> mice showed microvesicular steatosis, in contrast to the littermate controls (which had macrovesicular steatosis), suggesting decreased lipid accumulation in the liver of *K-ras*<sup>G12Dint</sup> mice. **f**, Messenger RNA expression levels of F4/80 and TNF- $\alpha$  were analysed by reverse transcriptase (RT)-PCR (HFD, LSL-*K-ras*<sup>G12D/+</sup> controls  $n = 3$ , *K-ras*<sup>G12Dint</sup> mice  $n = 6$ ). Plasma TNF- $\alpha$  levels determined by enzyme-linked immunosorbent assay (ELISA) showed decreased levels in *K-ras*<sup>G12Dint</sup> mice (HFD, LSL-*K-ras*<sup>G12D/+</sup> controls  $n = 7$ , *K-ras*<sup>G12Dint</sup> mice  $n = 11$ ).  $P$  values were determined by  $t$ -test. The error bars indicate s.e.m. \*,  $P \leq 0.05$ .





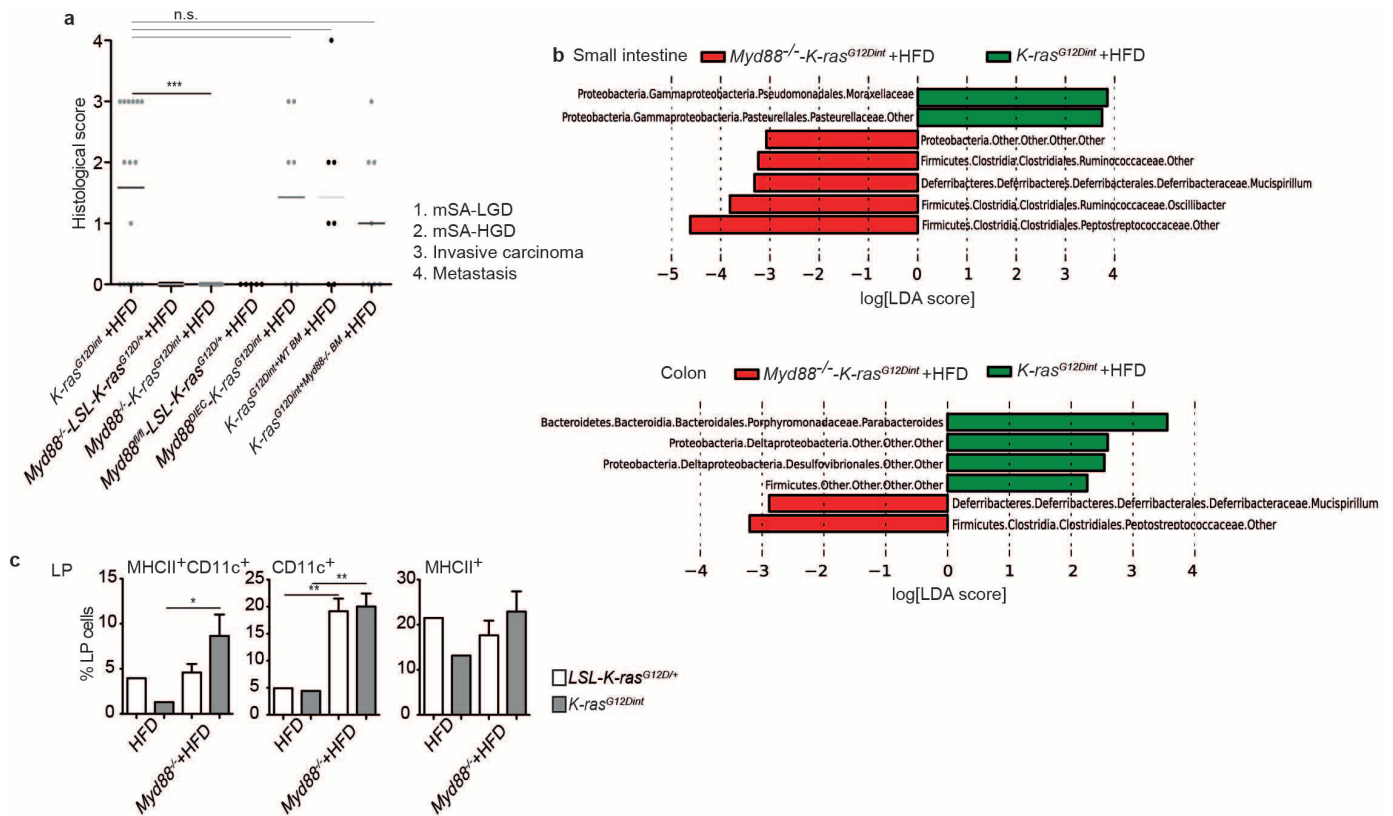
**Extended Data Figure 2 | The host immune response is dampened in *K-ras*<sup>G12Dint</sup> mice.** **a**, Relative mRNA expression levels of genes involved in the immune response were analysed by RT-PCR in duodenal samples from mice under the ND or the HFD regimen (ND, *LSL-K-ras*<sup>G12D/+</sup> controls *n* = 3, *K-ras*<sup>G12Dint</sup> mice *n* = 3; HFD, *LSL-K-ras*<sup>G12D/+</sup> controls *n* = 3, *K-ras*<sup>G12Dint</sup> mice *n* = 6). *P* values were determined by one-way analysis of variance (ANOVA) and adjusted for the number of comparisons by using the Bonferroni method. The error bars indicate s.e.m. \*\*, *P* ≤ 0.01. #, *P* ≤ 0.05 and ##, *P* ≤ 0.01 compared with littermate controls. **b**, Azure eosin staining of duodenal samples showed decreased amounts of granules with antimicrobial peptides in the crypts of *K-ras*<sup>G12Dint</sup> mice. **c**, The expression of differentiation markers for Paneth cells (cryptdin), epithelial cells (sucrase-isomaltase) and enteroendocrine cells (synaptophysin), as well as mucins in the duodenum, was

analysed by RT-PCR under the ND or the HFD regimen (ND, *LSL-K-ras*<sup>G12D/+</sup> controls *n* = 3, *K-ras*<sup>G12Dint</sup> mice *n* = 3; HFD, *LSL-K-ras*<sup>G12D/+</sup> controls *n* = 3, *K-ras*<sup>G12Dint</sup> mice *n* = 4). *P* values were determined by one-way ANOVA and adjusted for the number of comparisons by using the Bonferroni method. The error bars indicate s.e.m. \*, *P* ≤ 0.05. \*\*, *P* ≤ 0.01. #, *P* ≤ 0.05 compared with littermate controls. **d**, Flow cytometric analysis of cells from the lamina propria (LP) and Peyer's Patches (PP) indicated the presence of CD11c<sup>+</sup> dendritic cells (DCs) and the expression of MHC class II molecules in *K-ras*<sup>G12Dint</sup> mice and littermate controls on the ND or the HFD regimen (*n* = 2). *P* values were determined by one-way ANOVA and adjusted for the number of comparisons by using the Bonferroni method. The error bars indicate s.e.m. \*, *P* ≤ 0.05; \*\*, *P* ≤ 0.01. #, *P* ≤ 0.05 compared with littermate controls.



**Extended Data Figure 3 | An HFD leads to community changes in the gut microbiota.** **a, b,** Linear discriminant analysis (LDA) effect size (LEfSe) results showed bacteria that were significantly different in abundance among *K-ras*<sup>G12Dint</sup> mice and littermate controls on the HFD and indicated the effect size of each differentially abundant bacterial taxon in the small intestine

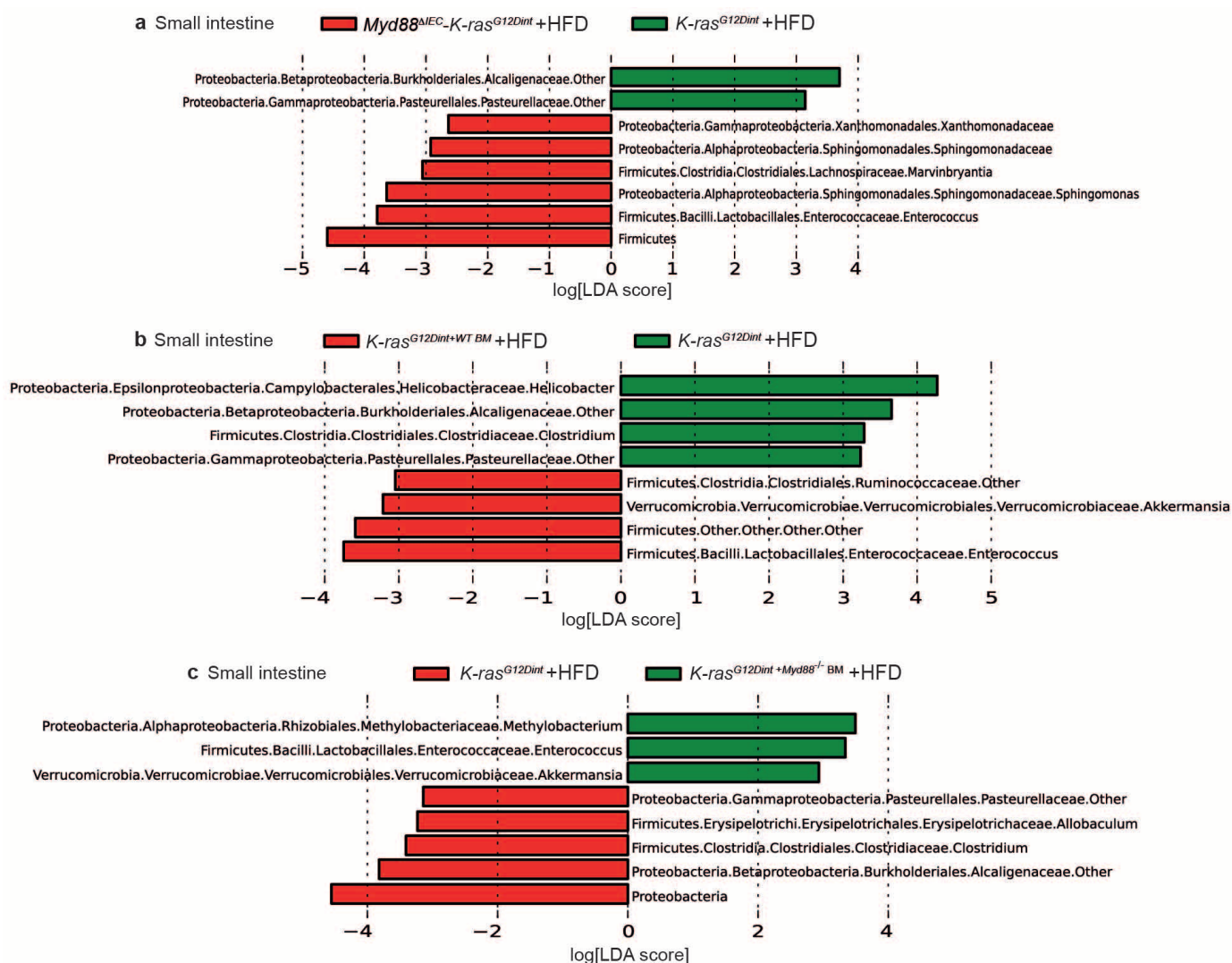
(*LSL-K-ras*<sup>G12D/+</sup> controls,  $n = 3$ ; *K-ras*<sup>G12Dint</sup> mice,  $n = 8$ ) (**a**) and colon (*LSL-K-ras*<sup>G12D/+</sup> controls,  $n = 5$ ; *K-ras*<sup>G12Dint</sup> mice,  $n = 7$ ) (**b**). **c,** The relative abundance of *Escherichia/Shigella* spp. and *Helicobacter* spp. in the small intestine of *K-ras*<sup>G12Dint</sup> mice on the ND or the HFD.



**Extended Data Figure 4 | Systemic deletion of *Myd88* prevents tumour progression in *K-ras*<sup>G12Dint</sup> mice.** **a**, Histological scores for the small intestine showed complete lack of tumour progression in HFD-fed *K-ras*<sup>G12Dint</sup> mice with systemic *Myd88* deletion. However, tissue-specific deletion of *Myd88* did not confer any protection against tumour progression in *K-ras*<sup>G12Dint</sup> mice (*Myd88*<sup>-/-</sup> LSL-*K-ras*<sup>G12D/+</sup> controls, *n* = 19; *Myd88*<sup>-/-</sup> *K-ras*<sup>G12Dint</sup> mice, *n* = 15; *Myd88*<sup>fl/fl</sup> LSL-*K-ras*<sup>G12D/+</sup> controls, *n* = 5; *Myd88*<sup>IEC</sup> *K-ras*<sup>G12Dint</sup> mice, *n* = 7; *K-ras*<sup>G12Dint</sup> + WT BM mice, *n* = 7; *K-ras*<sup>G12Dint</sup> + *Myd88*<sup>-/-</sup> BM mice, *n* = 8). Each point represents one animal, and the lines indicate means. A two-sided Fisher test was applied. Adjustments on pairwise *t*-tests were made using the single-step method. \*\*\*, *P* ≤ 0.001; NS, not significant. *Myd88*<sup>IEC</sup> *K-ras*<sup>G12Dint</sup>, *K-ras*<sup>G12Dint</sup> mice with IEC-specific deletion of *Myd88*; *K-ras*<sup>G12Dint</sup> + WT BM, *K-ras*<sup>G12Dint</sup> mice transplanted with wild-type bone marrow; *K-ras*<sup>G12Dint</sup> + *Myd88*<sup>-/-</sup> BM, *K-ras*<sup>G12Dint</sup> mice transplanted with

*Myd88*-deficient bone marrow. **b**, LEfSe results showed bacteria that were significantly different in abundance among HFD-fed *K-ras*<sup>G12Dint</sup> mice with (*n* = 8) or without (*n* = 4) systemic *Myd88* deletion. Peptostreptococcaceae, Deferribacteraceae and Ruminococcaceae in the small intestine, as well as Peptostreptococcaceae and Deferribacteraceae in the colon, became abundant after *Myd88* deficiency (*K-ras*<sup>G12Dint</sup> mice, *n* = 7; *Myd88*<sup>-/-</sup> *K-ras*<sup>G12Dint</sup> mice, *n* = 4). **c**, Flow cytometric analysis of LP cells indicated that the decreased recruitment of, and surface antigen presentation by, CD11c<sup>+</sup> DCs following the HFD was partially attenuated after *Myd88* deletion (*Myd88*<sup>-/-</sup> LSL-*K-ras*<sup>G12D/+</sup> controls, *n* = 8; *Myd88*<sup>-/-</sup> *K-ras*<sup>G12Dint</sup> mice, *n* = 4; LSL-*K-ras*<sup>G12D/+</sup> controls, *n* = 2; *K-ras*<sup>G12Dint</sup> mice, *n* = 2). *P* values were determined by one-way ANOVA and adjusted for the number of comparisons by using the Bonferroni method. The error bars indicate s.e.m. \*, *P* ≤ 0.05; \*\*, *P* ≤ 0.01.

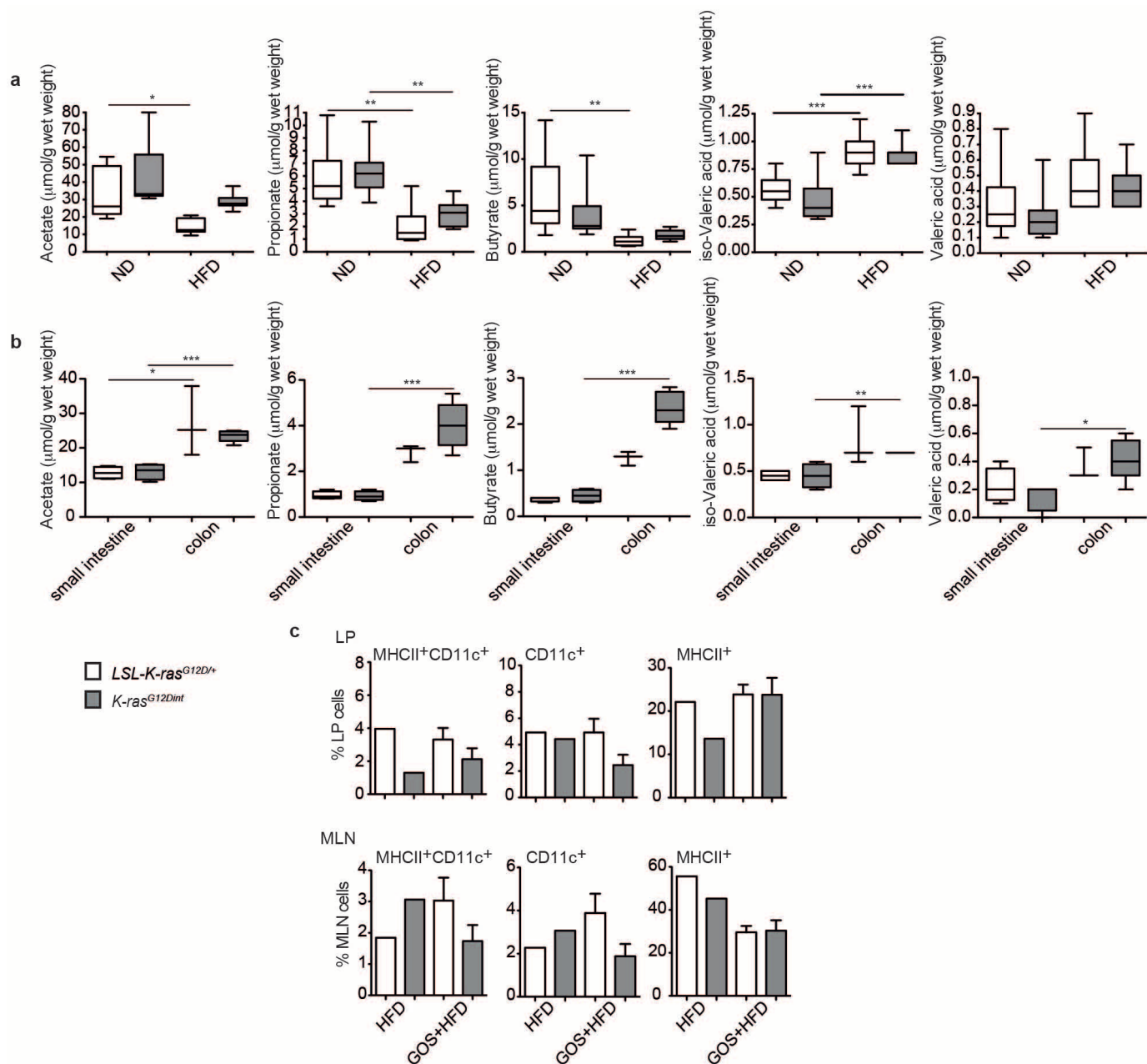




**Extended Data Figure 5 | Bacterial composition differs between *K-ras<sup>G12Dint</sup>* mice with tissue-specific and systemic deletion of *Myd88*.** LEfSe results showed that the composition of the small intestinal microbiota was significantly different between HFD-fed *K-ras<sup>G12Dint</sup>* mice with and without

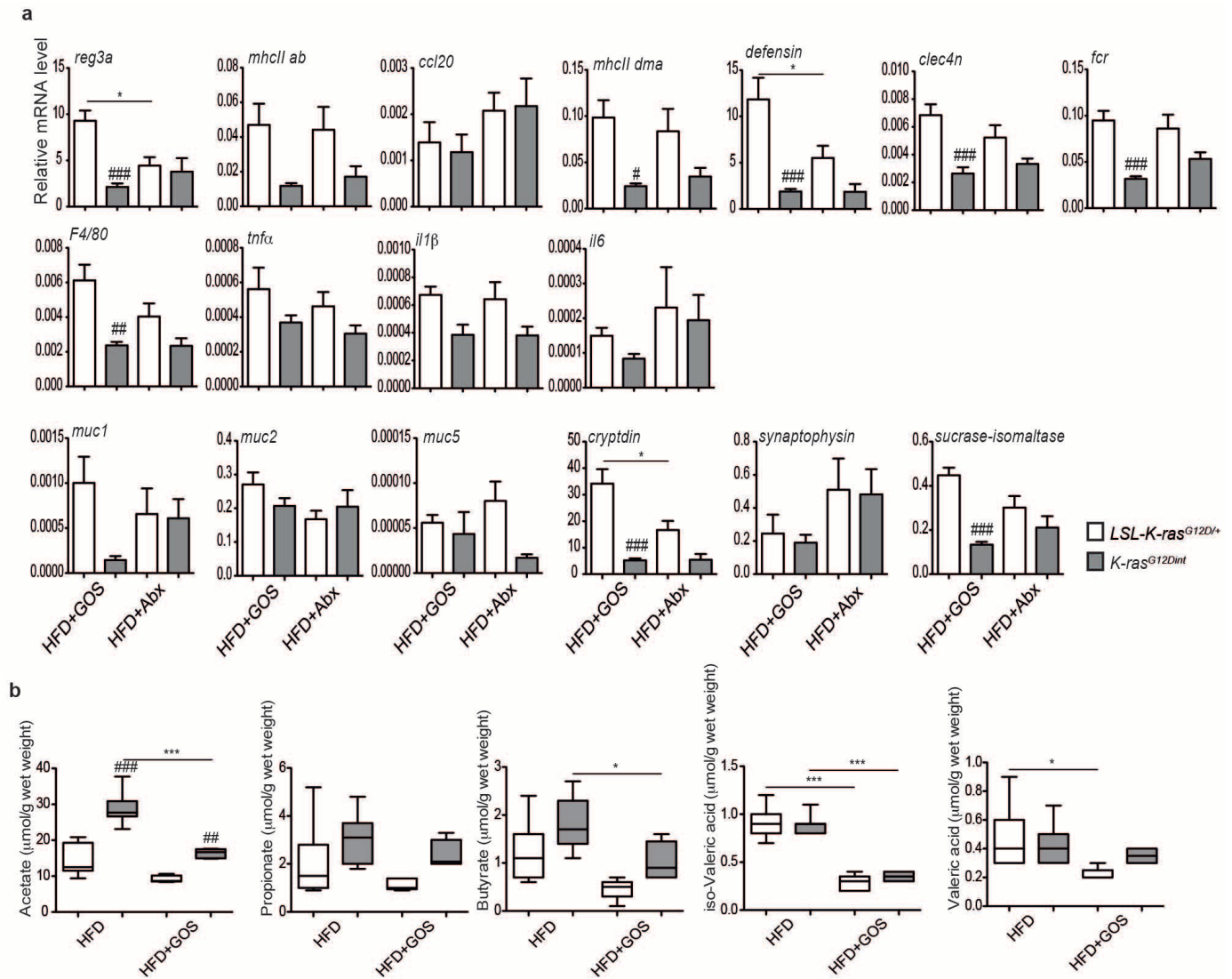
*Myd88* deletion in the IECs (a) or the haematopoietic cells (b, c) (*Myd88<sup>IEC</sup>* *K-ras<sup>G12Dint</sup>* mice, *n* = 4; *K-ras<sup>G12Dint</sup> + WT BM* mice, *n* = 4; *K-ras<sup>G12Dint</sup> + Myd88<sup>-/-</sup> BM* mice, *n* = 4; *K-ras<sup>G12Dint</sup>* mice, *n* = 8).





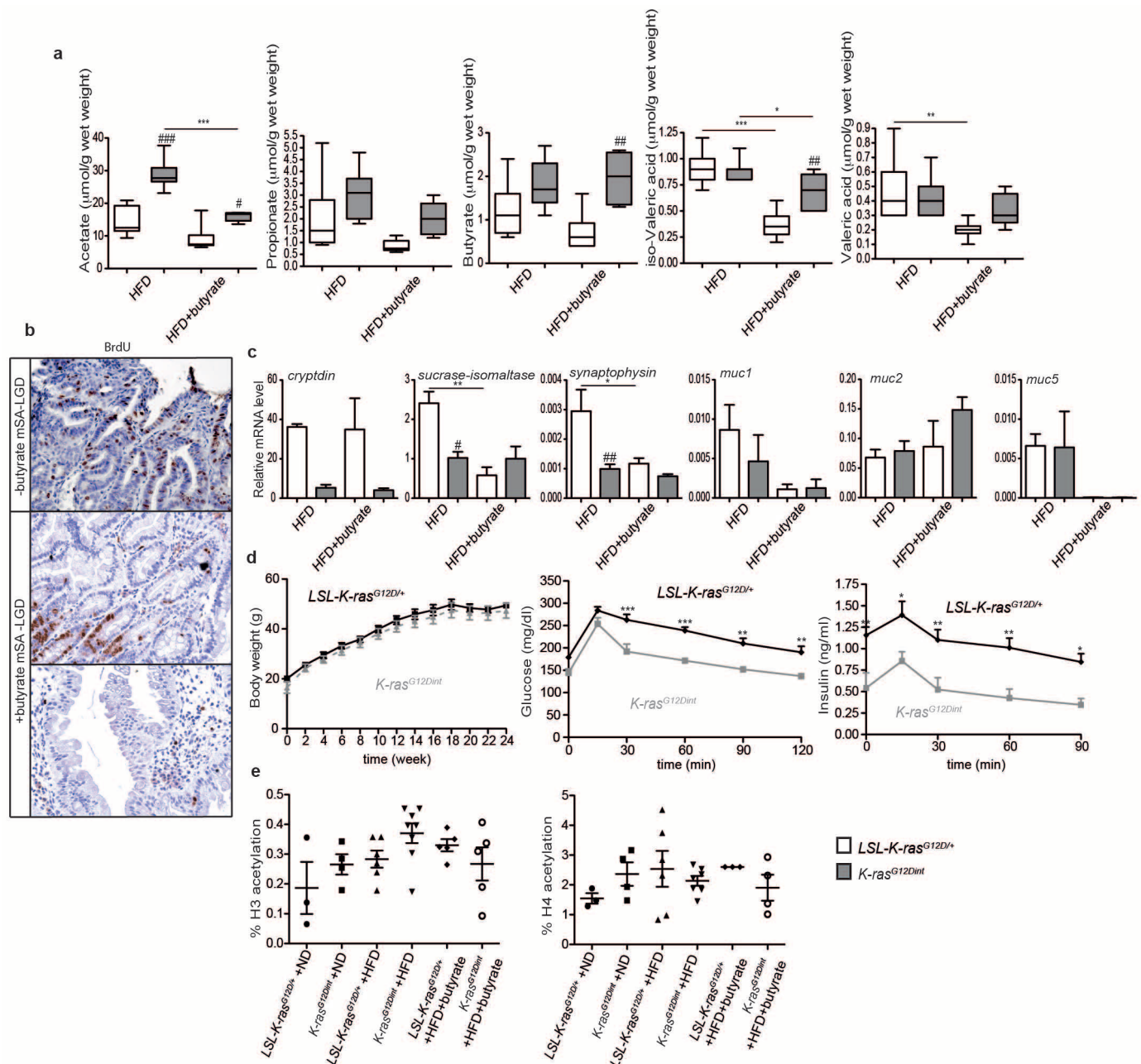
**Extended Data Figure 6 | An HFD decreases SCFA concentrations.** **a**, The HFD decreased the acetate, butyrate and propionate concentrations in stool samples from *K-ras*<sup>G12Dint</sup> mice and littermate controls, whereas the isovaleric acid and valeric acid concentrations increased (ND, *LSL-K-ras*<sup>G12D/+</sup> controls *n* = 6, *K-ras*<sup>G12Dint</sup> mice *n* = 8; HFD, *LSL-K-ras*<sup>G12D/+</sup> controls *n* = 7, *K-ras*<sup>G12Dint</sup> mice *n* = 11). *P* values were determined by one-way ANOVA and adjusted for the number of comparisons by using the Bonferroni method. The error bars indicate s.e.m. \*, *P* ≤ 0.05; \*\*, *P* ≤ 0.01; \*\*\*, *P* ≤ 0.001. **b**, SCFA concentrations of small intestinal samples (*LSL-K-ras*<sup>G12D/+</sup> controls, *n* = 4; *K-ras*<sup>G12Dint</sup> mice, *n* = 4) and colonic samples (*LSL-K-ras*<sup>G12D/+</sup> controls, *n* = 3; *K-ras*<sup>G12Dint</sup> mice, *n* = 5) from *K-ras*<sup>G12Dint</sup> and littermate controls on

the HFD supplemented with arabinogalactan. *P* values were determined by one-way ANOVA followed by Bonferroni's multiple comparison test. The error bars indicate s.e.m. \*, *P* ≤ 0.05; \*\*, *P* ≤ 0.01; \*\*\*, *P* ≤ 0.001. **c**, Flow cytometric analysis of cells from the LP and the MLNs revealed compromised DC recruitment and decreased surface antigen presentation in mice fed the HFD supplemented with GOS (HFD, *LSL-K-ras*<sup>G12D/+</sup> controls *n* = 2, *K-ras*<sup>G12Dint</sup> mice *n* = 2; HFD + GOS, *LSL-K-ras*<sup>G12D/+</sup> controls *n* = 5, *K-ras*<sup>G12Dint</sup> mice *n* = 3). *P* values were determined by one-way ANOVA and adjusted for the number of comparisons by using the Bonferroni method. The error bars indicate s.e.m. The differences are not significant.



**Extended Data Figure 7 | Prebiotic supplementation does not protect *K-ras<sup>G12Dint</sup>* mice against HFD-induced tumorigenesis.** **a**, Relative mRNA expression levels for genes involved in the immune response and those that encode mucins and differentiation markers for Paneth, enteroendocrine and epithelial cells in duodenal samples (HFD + GOS, *LSL-K-ras<sup>G12D/+</sup>* controls  $n = 5$ , *K-ras<sup>G12Dint</sup>* mice  $n = 6$ ; HFD + antibiotics (Abx), *LSL-K-ras<sup>G12D/+</sup>* controls  $n = 5$ , *K-ras<sup>G12Dint</sup>* mice  $n = 6$ ).  $P$  values were determined by one-way ANOVA and adjusted for the number of comparisons by using the Bonferroni method. The error bars indicate s.e.m. \*,  $P \leq 0.05$ , \*\*,  $P \leq 0.01$ , \*\*\*,  $P \leq 0.001$  compared with littermate controls.

and \*\*\*,  $P \leq 0.001$  compared with littermate controls. **b**, Prebiotic supplementation had little or no effect on stool SCFA concentrations (HFD, *LSL-K-ras<sup>G12D/+</sup>* controls  $n = 7$ , *K-ras<sup>G12Dint</sup>* mice  $n = 11$ ; HFD + GOS, *LSL-K-ras<sup>G12D/+</sup>* controls  $n = 5$ , *K-ras<sup>G12Dint</sup>* mice  $n = 6$ ).  $P$  values were determined by one-way ANOVA and adjusted for the number of comparisons by using the Bonferroni method. The error bars indicate s.e.m. \*,  $P \leq 0.05$ , \*\*\*,  $P \leq 0.001$ . #,  $P \leq 0.01$  and ###,  $P \leq 0.001$  compared with littermate controls.

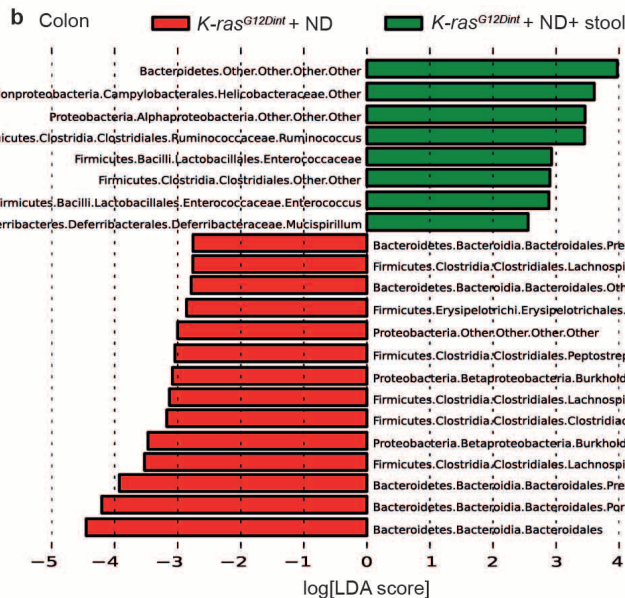
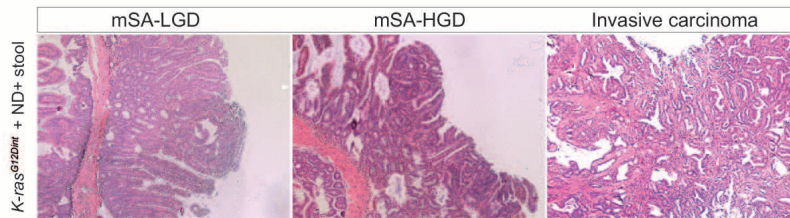


**Extended Data Figure 8 | Butyrate attenuates K-Ras signalling and has systemic effects on metabolic parameters.** **a**, Sodium butyrate treatment only slightly affected butyrate and propionate concentrations in the stool compared with prebiotic supplementation (HFD, *LSL-K-ras<sup>G12D/+</sup>* controls *n* = 7, *K-ras<sup>G12Dint</sup>* mice *n* = 11; HFD + butyrate, *LSL-K-ras<sup>G12D/+</sup>* controls *n* = 6, *K-ras<sup>G12Dint</sup>* mice *n* = 5). *P* values were determined by one-way ANOVA and adjusted for the number of comparisons by using the Bonferroni method. The error bars indicate s.e.m. \*, *P* ≤ 0.05; \*\*, *P* ≤ 0.01; \*\*\*, *P* ≤ 0.001. #, *P* ≤ 0.05, ##, *P* ≤ 0.01 and ###, *P* ≤ 0.001 compared with littermate controls. **b**, The higher proliferation levels observed in the duodenum of *K-ras<sup>G12Dint</sup>* mice were decreased following butyrate supplementation. **c**, The expression of differentiation markers and mucins in the duodenum (HFD, *LSL-K-ras<sup>G12D/+</sup>* controls *n* = 3, *K-ras<sup>G12Dint</sup>* mice *n* = 4; HFD + butyrate, *LSL-K-ras<sup>G12D/+</sup>*

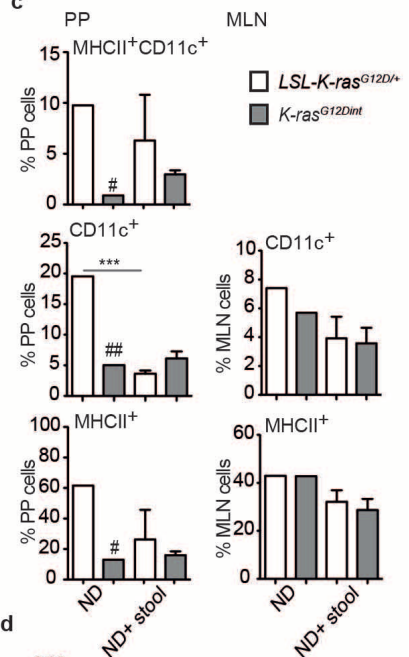
controls *n* = 3, *K-ras<sup>G12Dint</sup>* mice *n* = 3). *P* values were determined by one-way ANOVA and adjusted for the number of comparisons by using the Bonferroni method. The error bars indicate s.e.m. \*, *P* ≤ 0.05; \*\*, *P* ≤ 0.01. #, *P* ≤ 0.05 and ##, *P* ≤ 0.01 compared with littermate controls. **d**, Butyrate had systemic effects and protected *K-ras<sup>G12Dint</sup>* mice and littermate controls against HFD-induced hyperinsulinaemia (*n* = 5 per group). Data were assessed by *t*-test, and *P* values were adjusted for multiple testing. The error bars indicate s.e.m. \*, *P* ≤ 0.05; \*\*, *P* ≤ 0.01; \*\*\*, *P* ≤ 0.001. **e**, Butyrate treatment did not affect H3 or H4 acetylation (*n* = 3–8). The data were assessed by one-way ANOVA, and *P* values were adjusted for the number of comparisons by using the Bonferroni method. The error bars indicate s.e.m. The differences are not significant.



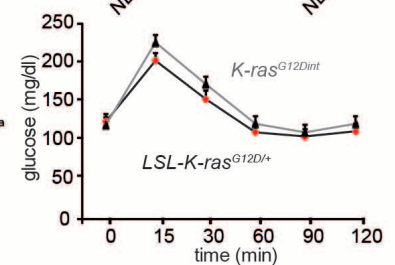
a



c



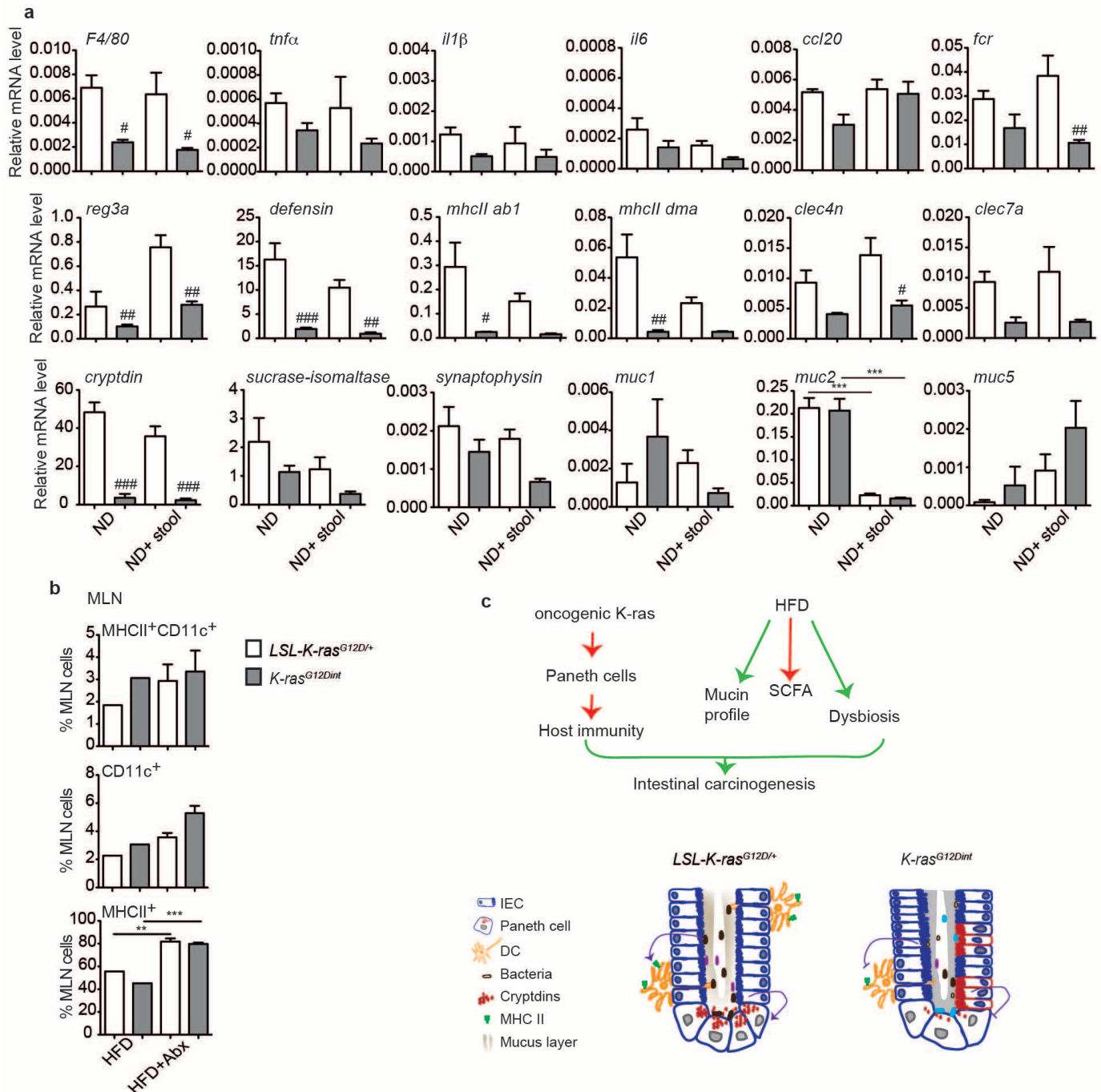
d



**Extended Data Figure 9 | HFD-induced dysbiosis and the associated cancer risk can be transferred to *K-ras*<sup>G12Dint</sup> mice on a normal diet.** **a**, Following 1 week of antibiotic cocktail treatment, *K-ras*<sup>G12Dint</sup> mice and littermate controls (approximately 7 weeks of age) on the ND regimen were gavaged three times a week with fresh stool pellets from HFD-fed mutants (which had been HFD fed for 24 weeks on the day of first transfer) for a total of 15 weeks. Haematoxylin and eosin staining of duodenal samples from three gavaged *K-ras*<sup>G12Dint</sup> mice show mSA-LGD and mSA-HGD, as well as invasive carcinoma development. **b**, LefSe results showed bacteria that were significantly different in abundance between ND-fed *K-ras*<sup>G12Dint</sup> mice that had been gavaged with stool samples from HFD-fed *K-ras*<sup>G12Dint</sup> donors and ND-fed *K-ras*<sup>G12Dint</sup> mice that had not been gavaged. Helicobacteraceae, Enterococcaceae and Deferribacteraceae became more abundant in the colon

after stool transfer (*K-ras*<sup>G12Dint</sup> mice + ND,  $n = 3$ ; *K-ras*<sup>G12Dint</sup> mice + ND + stool,  $n = 6$ ). **c**, Flow cytometric analysis of cells from the PP and MLNs showed antigen presentation by CD11c<sup>+</sup> DCs (ND, *LSL-K-ras*<sup>G12D/+</sup> controls  $n = 2$ , *K-ras*<sup>G12Dint</sup> mice  $n = 2$ ; ND + stool, *LSL-K-ras*<sup>G12D/+</sup> controls  $n = 7$ , *K-ras*<sup>G12Dint</sup> mice  $n = 10$ ).  $P$  values were determined by one-way ANOVA and adjusted for the number of comparisons by using the Bonferroni method. The error bars indicate s.e.m. \*\*\*,  $P \leq 0.001$ . #,  $P \leq 0.05$  and ##,  $P \leq 0.01$  compared with littermate controls. **d**, Glucose clearance during a GTT in ND-fed *K-ras*<sup>G12Dint</sup> mice and littermate controls ( $n = 5$  per group) that had received stool samples from HFD-fed *K-ras*<sup>G12Dint</sup> donors. The results were analysed by  $t$ -test.  $P$  values were adjusted for multiple testing. The error bars indicate s.e.m. The differences are not significant.





**Extended Data Figure 10 | Antibiotic treatment blocks HFD-induced tumorigenesis in *K-ras*<sup>G12Dint</sup> mice.** **a**, The expression profiles of selected genes involved in antigen recognition, immune response, immune cell recruitment, differentiation markers and mucins in duodenal samples from *LSL-K-ras*<sup>G12D/+</sup> and *K-ras*<sup>G12Dint</sup> littermate animals (ND, *LSL-K-ras*<sup>G12D/+</sup> controls  $n = 3$ , *K-ras*<sup>G12Dint</sup> mice  $n = 3$ ; ND + stool, *LSL-K-ras*<sup>G12D/+</sup> controls  $n = 3$ , *K-ras*<sup>G12Dint</sup> mice  $n = 5$ ).  $P$  values were determined by one-way ANOVA and adjusted for the number of comparisons with the Bonferroni method. The error bars indicate s.e.m. \*\*\*,  $P \leq 0.001$ . #,  $P \leq 0.05$ , ##,  $P \leq 0.01$  and ###,  $P \leq 0.001$  compared with littermate controls. **b**, Fluorescence-activated cell

sorting (FACS) analysis of MLN cells indicates recruitment of, and antigen presentation by, CD11c<sup>+</sup> dendritic cells following treatment with antibiotics (HFD, *LSL-K-ras*<sup>G12D/+</sup> controls  $n = 2$ , *K-ras*<sup>G12Dint</sup> mice  $n = 2$ ; HFD + Abx, *LSL-K-ras*<sup>G12D/+</sup> controls  $n = 5$ , *K-ras*<sup>G12Dint</sup> mice  $n = 7$ ).  $P$  values were determined by one-way ANOVA and adjusted for the number of comparisons with the Bonferroni method. The error bars indicate s.e.m. \*\*\*,  $P \leq 0.001$ . The results are not significant. **c**, The mechanistic scheme suggests that HFD-induced changes in the bacterial community, SCFA levels and mucin profiles cooperate with an oncogene-associated decrease in host immunity, collectively enhancing carcinogenesis in the small intestine.

# Contrasting roles of histone 3 lysine 27 demethylases in acute lymphoblastic leukaemia

Panagiotis Ntziachristos<sup>1,2\*</sup>, Aristotelis Tsirigos<sup>1,3\*</sup>, G. Grant Welstead<sup>4,5†</sup>, Thomas Trimarchi<sup>1,2</sup>, Sofia Bakogianni<sup>1,2</sup>, Luyao Xu<sup>6</sup>, Evangelia Loizou<sup>1,2</sup>, Linda Holmfeldt<sup>7</sup>, Alexandros Strikoudis<sup>1,2</sup>, Bryan King<sup>1,2</sup>, Jasper Mullenders<sup>1,2</sup>, Jared Becksfort<sup>8</sup>, Jelena Nedjic<sup>1,2</sup>, Elisabeth Paietta<sup>9</sup>, Martin S. Tallman<sup>10</sup>, Jacob M. Rowe<sup>11,12</sup>, Giovanni Tonon<sup>13</sup>, Takashi Satoh<sup>14,15</sup>, Laurens Kruidenier<sup>16</sup>, Rab Prinjha<sup>16</sup>, Shizuo Akira<sup>14,15</sup>, Pieter Van Vlierberghe<sup>6,17</sup>, Adolfo A. Ferrando<sup>6,18,19</sup>, Rudolf Jaenisch<sup>4,5</sup>, Charles G. Mullighan<sup>7</sup> & Iannis Aifantis<sup>1,2</sup>

**T-cell acute lymphoblastic leukaemia (T-ALL) is a haematological malignancy with a dismal overall prognosis, including a relapse rate of up to 25%, mainly because of the lack of non-cytotoxic targeted therapy options. Drugs that target the function of key epigenetic factors have been approved in the context of haematopoietic disorders<sup>1</sup>, and mutations that affect chromatin modulators in a variety of leukaemias have recently been identified<sup>2,3</sup>; however, ‘epigenetic’ drugs are not currently used for T-ALL treatment. Recently, we described that the polycomb repressive complex 2 (PRC2) has a tumour-suppressor role in T-ALL<sup>4</sup>. Here we delineated the role of the histone 3 lysine 27 (H3K27) demethylases JMJD3 and UTX in T-ALL. We show that JMJD3 is essential for the initiation and maintenance of T-ALL, as it controls important oncogenic gene targets by modulating H3K27 methylation. By contrast, we found that UTX functions as a tumour suppressor and is frequently genetically inactivated in T-ALL. Moreover, we demonstrated that the small molecule inhibitor GSKJ4 (ref. 5) affects T-ALL growth, by targeting JMJD3 activity. These findings show that two proteins with a similar enzymatic function can have opposing roles in the context of the same disease, paving the way for treating haematopoietic malignancies with a new category of epigenetic inhibitors.**

In recent studies, we and other researchers have revealed that PRC2 has a key tumour-suppressor function, catalysing the methylation of H3K27 (refs 2, 4, 6). Since net H3K27me3 levels are dictated by the balance between histone methylation and active histone demethylation, we hypothesized that the removal of methyl groups from H3K27 is also an important process in T-ALL progression. We therefore investigated the possible roles of H3K27 demethylases in T-ALL (see Supplementary Notes for an extended introduction). Ubiquitously transcribed tetratricopeptide repeat X-linked protein (UTX)<sup>7,8</sup> (also known as KDM6A) is a ubiquitously expressed protein that controls the basal levels of H3K27me3 and the induction of ectoderm and mesoderm differentiation<sup>9,10</sup> and is essential for somatic cell reprogramming<sup>11</sup>. Jumonji D3 (JMJD3)<sup>7,8</sup> (also known as KDM6B) is induced upon inflammation<sup>12</sup> or exposure to viral and oncogenic stimuli<sup>13,14</sup>, and it controls neuronal and epidermal differentiation<sup>15,16</sup> and inhibits reprogramming<sup>17</sup>. UTX is a tumour suppressor in several solid tumours<sup>3,18–21</sup>. However, the roles of these two demethylases as direct modulators of the oncogenic state are largely uncharacterized<sup>13,14</sup>.

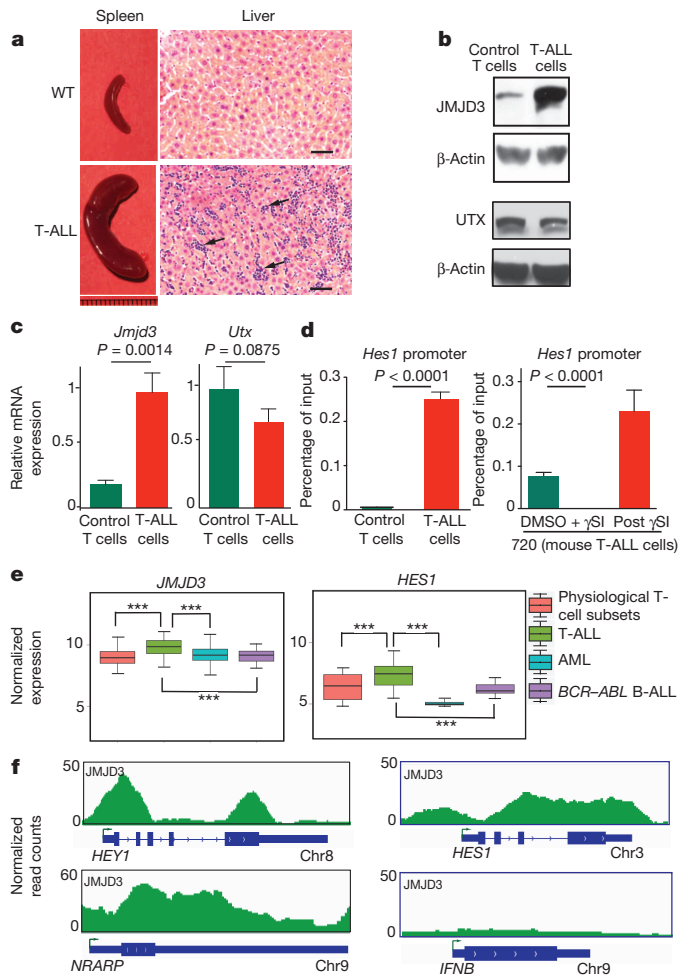
We have generated and studied NOTCH1-induced T-ALL animal models<sup>4</sup> (Fig. 1a), because activating mutations of NOTCH1 are a defining feature of T-ALL<sup>22</sup>. *Jmjd3* messenger RNA and protein expression levels were significantly higher in leukaemic cells than in untransformed CD4<sup>+</sup>CD8<sup>+</sup> (double positive) control T cells, which exhibit low levels of active NOTCH1, whereas *Utx* (and *Ezh2*)<sup>4</sup> expression did not change significantly (Fig. 1b, c and Supplementary Table 1) upon transformation. It has previously been shown that the transcription factor nuclear factor- $\kappa$ B (NF- $\kappa$ B) controls *JMJD3* expression during inflammation<sup>12</sup> and that NOTCH1 induces the NF- $\kappa$ B pathway in T-ALL<sup>23</sup>. Here we showed increased expression of the p65 subunit (also known as RELA) of NF- $\kappa$ B and its—but not NOTCH1—binding to *Jmjd3* control elements in mouse T-ALL cells (Extended Data Fig. 1a, b). Modulation of the levels of intracellular NOTCH1 or the activity of the NF- $\kappa$ B pathway significantly decreased the amount of NF- $\kappa$ B bound to the *Jmjd3* elements, as well as *Jmjd3* mRNA expression (Extended Data Fig. 1b–f). We then probed for JMJD3 binding to specific oncogenic loci, which has previously been shown to be important in T-ALL<sup>4</sup>. We found that JMJD3 binding was highly enriched on the *Hes1* promoter (Fig. 1d, left), and this binding depended on the activation of the NOTCH1 pathway and negatively correlated with the H3K27me3 levels (Extended Data Fig. 1g, h).

Analyses of human leukaemia cases<sup>2,24–26</sup> have shown that JMJD3 is more highly expressed in T-ALL cells than in normal T-cell progenitors<sup>24</sup> or in other types of leukaemia<sup>25,26</sup>, which is similar to the expression of the classic NOTCH1 target *HES1* (Fig. 1e). Genes that are co-expressed with *JMJD3* in human primary samples were found to exhibit loss of H3K27me3 during leukaemia progression (Extended Data Fig. 1i), suggesting a connection between the expression of JMJD3 and the H3K27me3 levels on specific targets.

Chromatin immunoprecipitation followed by sequencing (ChIP-seq) studies in human T-ALL cells (the cell line CUTTL1) showed that JMJD3 was bound to important NOTCH1 targets with oncogenic functions (such as *HEY1*, *NRARP* and *HES1*) (Fig. 1f). There was a significant co-occupancy of JMJD3 with NOTCH1 (ref. 27) (33% of the top JMJD3 peaks were occupied by NOTCH1, a 6.9-fold enrichment over control,  $P < 1 \times 10^{-3}$ ), the NOTCH1 partner RBP-J $\kappa$  and the activating mark H3K4me3 (ref. 27) (Extended Data Fig. 1j). The majority of JMJD3 binding sites were localized around the transcription start sites (TSSs) of

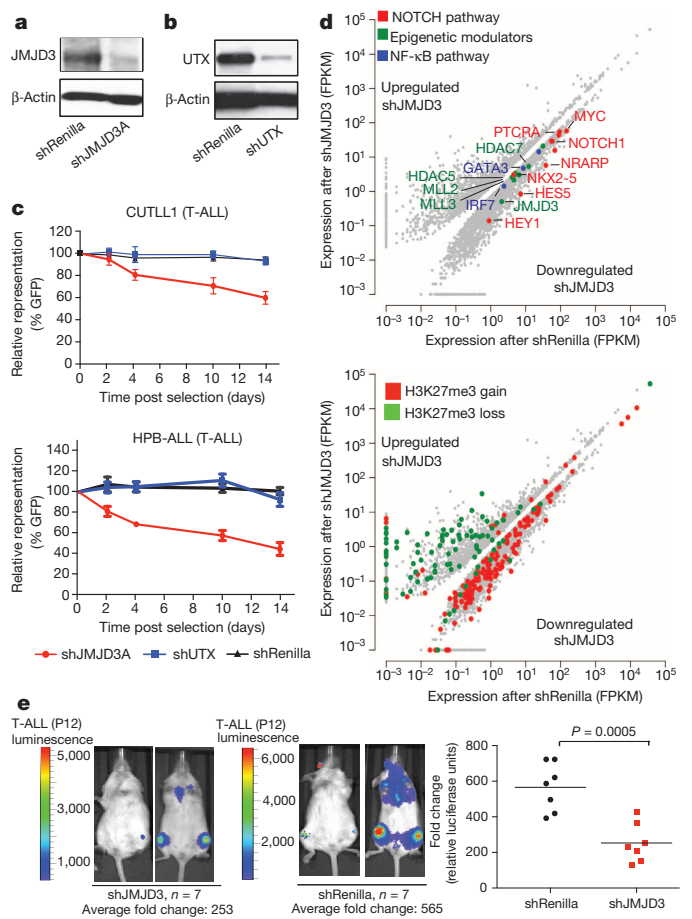
<sup>1</sup>Howard Hughes Medical Institute and Department of Pathology, NYU School of Medicine, New York, New York 10016, USA. <sup>2</sup>NYU Cancer Institute and Helen L. and Martin S. Kimmel Center for Stem Cell Biology, NYU School of Medicine, New York, New York 10016, USA. <sup>3</sup>Center for Health Informatics and Bioinformatics, NYU School of Medicine, New York, New York 10016, USA. <sup>4</sup>Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142, USA. <sup>5</sup>Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. <sup>6</sup>Institute for Cancer Genetics, Columbia University Medical Center, New York, New York 10032, USA. <sup>7</sup>Department of Pathology, St. Jude Children's Research Hospital, Memphis, Tennessee 38105, USA. <sup>8</sup>Department of Computational Biology, St. Jude Children's Research Hospital, Memphis, Tennessee 38105, USA. <sup>9</sup>Montefiore Medical Center North, Bronx, New York, New York 10467, USA. <sup>10</sup>Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA. <sup>11</sup>Technion, Israel Institute of Technology, Haifa 31096, Israel. <sup>12</sup>Shaare Zedek Medical Center, Jerusalem 9103102, Israel. <sup>13</sup>Functional Genomics of Cancer Unit, Division of Molecular Oncology, Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS) San Raffaele Scientific Institute, 20132 Milan, Italy. <sup>14</sup>Laboratory of Host Defense, WPI Immunology Frontier Research Center (WPI IFRC), Osaka University, 3-1 Yamada-oka, Suita, Osaka 565-0871, Japan. <sup>15</sup>Department of Host Defense, Research Institute for Microbial Diseases (RIMD), Osaka University, 3-1 Yamada-oka, Suita, Osaka 565-0871, Japan. <sup>16</sup>Epinova DPU, Immuno-Inflammation Therapy Area, GlaxoSmithKline R&D, Medicines Research Centre, Gunnels Wood Road, Stevenage SG1 2NY, UK. <sup>17</sup>Center for Medical Genetics, Ghent University Hospital, 9000 Ghent, Belgium. <sup>18</sup>Department of Pathology, Columbia University Medical Center, New York, New York 10032, USA. <sup>19</sup>Department of Pediatrics, Columbia University Medical Center, New York, New York 10032, USA. †Present address: Editas Medicine, Cambridge, Massachusetts 02142, USA.

\*These authors contributed equally to this work.



**Figure 1 | JMJD3 is highly expressed in T-ALL and controls the expression of important oncogenic targets.** **a**, Size comparison of the spleens (left) and haematoxylin and eosin staining of the liver (centre) of healthy (WT, top) and leukaemic (T-ALL, bottom) mice. The arrows denote leukaemic infiltration of the liver of the T-ALL mouse. Scale bar, 50  $\mu$ m. Representative samples from  $n = 3$  mice are shown. **b**, **c**, Protein (**b**) and transcript (**c**) levels of the demethylases JMJD3 and UTX in control T cells (CD4<sup>+</sup>CD8<sup>+</sup> (double positive) thymocytes) and mouse T-ALL cells. Representative samples (**b**) or the mean  $\pm$  s.d. of three mice is shown; values were normalized according to the sample with the highest expression value. **d**, ChIP for JMJD3 on the *Hes1* promoter in control T cells and mouse T-ALL cells (left) and upon  $\gamma$ -secretase inhibitor ( $\gamma$ SI) treatment of T-ALL cells (right) ( $n = 3$ ): data are shown as mean  $\pm$  s.d. DMSO, dimethylsulphoxide. **e**, Expression analysis of *JMJD3* and *HES1* among samples of acute T-cell leukaemia (T-ALL; 83 samples), acute B-cell leukaemia (B-ALL; 23) and acute myeloid leukaemia (AML; 537), as well as physiological T-cell subsets (24)<sup>24</sup> (quantile normalization across samples, see Methods). The data are shown as mean  $\pm$  s.d. The  $P$  values (Wilcoxon test) are as follows: for *JMJD3*, T-ALL versus physiological T cells,  $4.0 \times 10^{-6}$ ; T-ALL versus AML,  $1.1 \times 10^{-13}$ ; T-ALL versus B-ALL,  $2.2 \times 10^{-5}$ ; and for *HES1*, T-ALL versus physiological T cells,  $3.7 \times 10^{-4}$ ; T-ALL versus AML,  $3.5 \times 10^{-43}$ ; T-ALL versus B-ALL,  $1.3 \times 10^{-6}$ . \*\*\*, significant. **f**, Snapshots of JMJD3 binding in human T-ALL. Three NOTCH1 targets and the interferon- $\beta$  (*IFNB*) gene (negative control) are shown. Chr, chromosome.

genes (Extended Data Fig. 1k) in a fashion similar to NOTCH1 binding sites<sup>27</sup>. These results suggest a key role for JMJD3 in oncogenic programs in T-ALL, through interaction with NOTCH1. Protein immunoprecipitation studies in 293T cells (human embryonic kidney cells), as well as in mouse T-ALL cell lines, showed that JMJD3 is part of the NOTCH1 transcriptional complex, as it interacts directly with NOTCH1 and MAML1 (Extended Data Fig. 2a–c). By contrast, there was no NOTCH1 interaction with EZH2 or UTX. As JMJD3 has been shown to be a

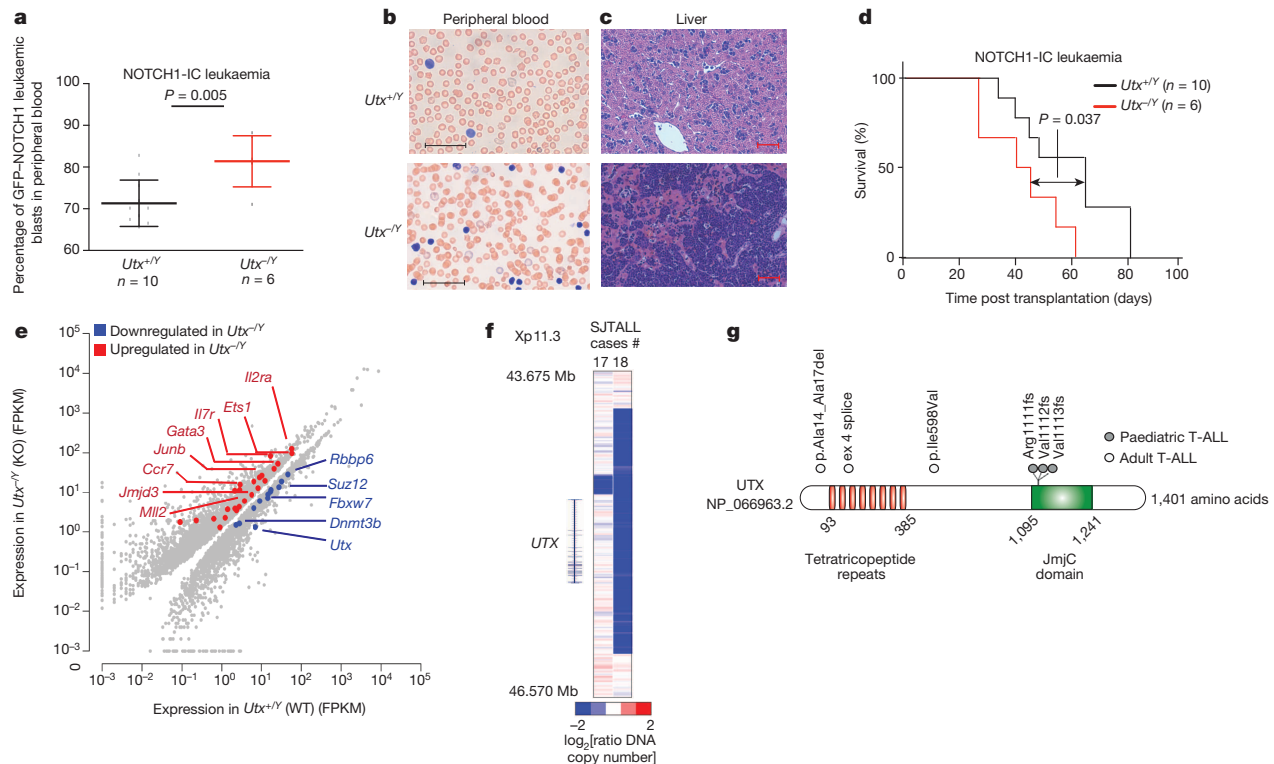


**Figure 2 | Dissecting the oncogenic role of JMJD3 in T-ALL.** **a**, **b**, The protein levels of JMJD3 (**a**) and UTX (**b**) in human T-ALL cells (CUTLL1) expressing the corresponding shRNAs against the two demethylases. Representative blots from three independent studies (biological replicates) are shown. **c**, Effects on human T-ALL cell proliferation of shRNA treatment, as measured by loss of green fluorescent protein (GFP)-expressing shRNA. For all cell lines, the mean  $\pm$  s.d. from three representative studies is shown. **d**, Differential expression analysis upon knockdown of *JMJD3* in T-ALL (top). The loci of the downregulated genes exhibit an increase in H3K27me3 (red dots, bottom), whereas the upregulated genes exhibit a decrease in H3K27me3. The data shown are representative of three independent studies. FPKM, fragments per kilobase of transcript per million fragments mapped. **e**, *In vivo* growth of P12 T-ALL cells in intravenous xenograft studies upon genomic ablation of *JMJD3* (left) and with a Renilla control (centre). One million P12 cells were injected into each of seven animals. Sublethally irradiated NRG (immunocompromised) mice were used as recipients, and transplanted leukaemic cell growth was compared with the baseline (day 0). Day 0 was set as the first day when substantially detectable luciferase intensity was measured. The last day of the experiment was the day that either the luciferase intensity reached saturation or the mice were killed for humane reasons. Horizontal bars, means.

member of MLL complexes<sup>12</sup>, we tested whether JMJD3 interacted with WDR5, a key subunit of the MLL complex. We found that JMJD3 interacted with WDR5 (Extended Data Fig. 2b), suggesting a potential NOTCH1–JMJD3–MLL complex on target promoters.

To clarify the role of JMJD3 and UTX in the maintenance of leukaemia, we performed genomic knockdown of *JMJD3* in human T-ALL cells using two short hairpin RNAs (shRNAs) (Fig. 2a, b and Extended Data Fig. 2d). Treatment with shJMJD3 but not shUTX affected the viability of leukaemic cells, as shown in loss of representation studies and apoptosis assays, and this finding is in contrast to the viability of myeloid leukaemia lines used as controls (Fig. 2c and Extended Data Fig. 2e, f). The expression of NOTCH1 targets was negatively affected by shJMJD3,





**Figure 3 | The demethylase UTX acts as a tumour suppressor in T-ALL.** **a–c**, Monitoring the initiation and progression of T-cell leukaemia in a NOTCH1-overexpressing model of T-ALL. Leukaemic blasts (expressed as NOTCH1-IC-GFP-positive cells) in the peripheral blood (**a**, mean  $\pm$  s.d.) and in a blood smear (**b**) and leukaemic cell infiltration of the liver (**c**) of male wild-type ( $Utx^{+/Y}$ ,  $n = 10$ ) and knockout ( $Utx^{-/Y}$ ,  $n = 6$ ) mice are shown. NOTCH1-IC, intracellular part of NOTCH1. **d**, Survival studies of mice transplanted with haematopoietic progenitors from the wild-type ( $Utx^{+/Y}$ ,  $n = 10$ ) and knockout ( $Utx^{-/Y}$ ,  $n = 6$ ) backgrounds expressing NOTCH1-IC. **e**, Scatter plot summarizing the major genome-wide expression differences between T-ALL tumours of the wild-type ( $Utx^{+/Y}$ ) and knockout ( $Utx^{-/Y}$ )

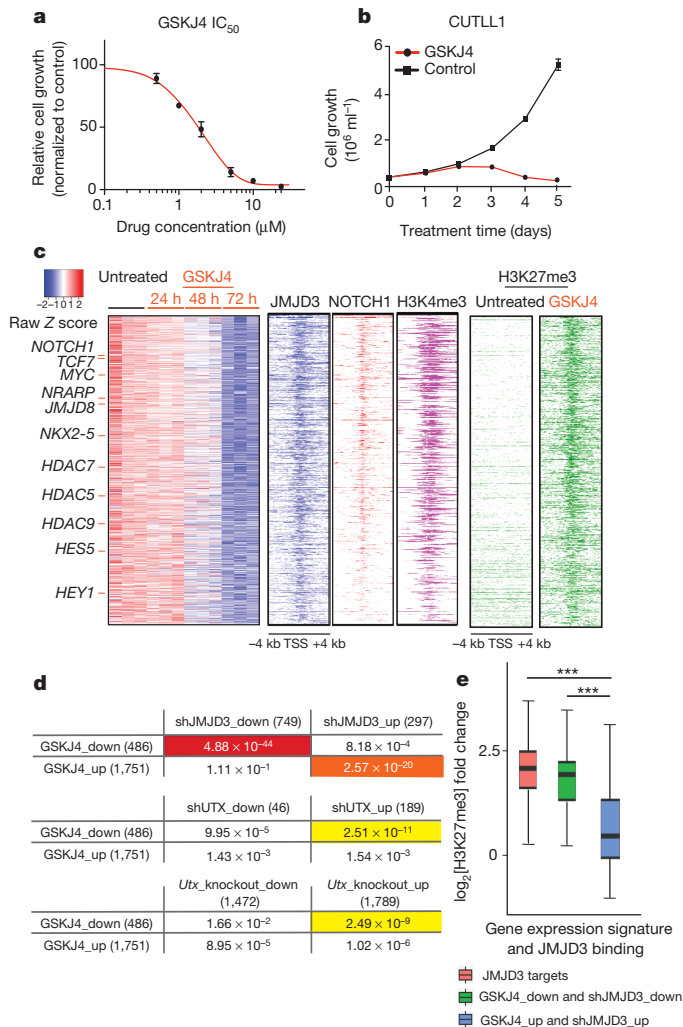
backgrounds. RNA sequencing was performed using three pairs of wild-type and *Utx* knockout (KO) NOTCH1-IC tumours (spleen and bone marrow). **f, g**, Analysis of genetic status of the *UTX* (*KDM6A*) locus in paediatric T cell leukaemia ( $n = 107$ ). Affymetrix SNP6.0 microarrays (**f**) for assessing genomic deletions. Illustration of the human *UTX* protein (**g**) depicting three frameshift (**fs**) mutations in paediatric T-ALL (grey circles), as well as one in-frame deletion (p.Ala14\_Ala17del), one splice acceptor site mutation (exon (ex) 4 splice) and one missense mutation (p.Ile598Val) in adult T-ALL (white circles), as identified by targeted Sanger sequencing. The jumonji domain (JmjC) and the tetratricopeptide repeats are shown. SJTALL, St. Jude Children's Research Hospital sample depository of T-ALL samples.

and this was accompanied by loss of JMJD3 and gain of H3K27me3 on their promoters (Extended Data Fig. 3a–e). Genome-wide expression analysis showed that more transcripts were significantly downregulated by shJMJD3 treatment than were upregulated (749 protein-coding genes versus 297; Fig. 2d, top, and Extended Data Fig. 3f), in agreement with the role of JMJD3 as a transcriptional activator. The downregulated genes were found to be significantly enriched in genes that gained H3K27me3 on their promoters (Fig. 2d, bottom;  $P = 1.02 \times 10^{-7}$ ). The shUTX-downregulated and shUTX-upregulated gene signatures were reversed in terms of the gene numbers (46 downregulated and 189 upregulated protein-coding genes, compared with both shRenilla (control) and shJMJD3). Intriguingly, *JMJD3* expression itself was significantly upregulated upon *UTX* silencing (Extended Data Fig. 3a). Well-characterized NOTCH1 targets, as well as genes in the NF- $\kappa$ B pathway were downregulated as part of the *JMJD3* signature (Fig. 2d, top, and Extended Data Fig. 3g). These findings were confirmed using additional human T-ALL cell lines with high levels of oncogenic NOTCH1 activity<sup>22</sup> (Extended Data Fig. 3h, i). Subcutaneous or intravenous xenograft models of T-ALL cell lines (CUTLL1, CEM and P12) treated with either of the two shRNAs against *JMJD3* (shJMJD3A and shJMJD3B) and transplanted into immunocompromised mice (NRG mice; NOD *Rag1*<sup>-/-</sup> *Il2rg*<sup>-/-</sup>) showed a significant growth disadvantage compared with shRenilla-treated cell lines (Fig. 2e and Extended Data Fig. 4a–f). Interestingly, silencing of *UTX* led to enhanced proliferation in many cases, suggesting a possible tumour-suppressor function *in vivo* (Extended Data Fig. 4g).

To examine the potential roles of *UTX* and *JMJD3* in the induction of T-ALL, we performed bone marrow transplantation experiments using haematopoietic stem cells from *Utx* and *Jmjd3* germline knockout mice. Although female *Utx*<sup>-/-</sup> mice die at E9.5 because of defects in mesoderm development, a small fraction of male *Utx*<sup>-/Y</sup> mice survive to adulthood as a result of compensation by *UTY*<sup>28</sup>. Despite T-cell development being largely unaffected (Extended Data Fig. 5a, b), T-ALL kinetics were significantly faster on the *Utx*<sup>-/Y</sup> background, as determined by leukaemic burden quantification in the peripheral blood and infiltration of the spleen (data not shown) and liver (Fig. 3a–c and Extended Data Fig. 5c–e). Moreover, mice succumbed to the disease with a significantly shorter latency in the absence of *Utx* than in the *Utx*<sup>+/Y</sup> and *Utx*<sup>+/+</sup> genotypes (Fig. 3d and Extended Data Fig. 5f–h). These experiments provide the first *in vivo* analysis of the tumour-suppressor role of *UTX* in any tumour type.

To delineate the potential mechanism underlying *UTX* action, we analysed the gene expression of sorted leukaemic blasts from the spleen or bone marrow of wild-type (*Utx*<sup>+/Y</sup>/*Utx*<sup>+/+</sup>) or knockout (*Utx*<sup>-/Y</sup>) mice (Fig. 3e). This analysis showed that *UTX* positively controls important tumour-suppressor genes, such as retinoblastoma binding protein 6 (*Rbbp6*), the inhibitor of NOTCH1 pathway activity *Fbxw7* and the PRC2 member *Suz12*; by contrast, genes with an oncogenic role in T-ALL, including *Jmjd3*, were upregulated (Fig. 3e and Extended Data Fig. 5i). These studies strongly suggested that *UTX* might act as a tumour suppressor in human T-ALL. We thus screened a panel of primary paediatric T-ALL samples<sup>2</sup> for genetic alterations of the *UTX* locus. Analysis





**Figure 4 | Pharmacological targeting of T-ALL through specific inhibition of the demethylase activity of JMJD3.** **a**, Dose-dependent effect of the inhibitor GSKJ4 (normalized to a control inhibitor, GSKJ5)<sup>2</sup> on CUTLL1 cell proliferation. The data are shown as mean  $\pm$  s.d. **b**, Effect of GSKJ4 (at 2 μM) on CUTLL1 T-ALL cells. The data are shown as mean  $\pm$  s.d. **c**, Heatmap representation of GSKJ4-associated changes in gene expression (left, three biological replicates) and H3K27me3 changes (right) of 486 significantly downregulated coding transcripts in CUTLL1 T-ALL cells over a period of 72 h. Centre, Occupancy by JMJD3 and NOTCH1 and H3K4me3 marks in respective 4-kilobase (kb)-flanked TSS regions (indicated in different colours from each other). **d**, Comparison of the shJMJD3 and shUTX expression-based signatures with the GSKJ4-induced expression changes. Note the highly significant overlap between the genes downregulated by shJMJD3 and GSKJ4. Different colours represent different levels of significance: red, high significance; orange, medium significance; and yellow, low significance. **e**, Box plots representing mean  $\pm$  s.d. GSKJ4-induced H3K27me3 changes in JMJD3 target genes, as well as in the commonly downregulated genes in shJMJD3- and GSKJ4-treated cells. Genes upregulated by both shJMJD3 and GSKJ4 treatments were used as the negative controls. The *P* values are as follows: JMJD3 targets versus GSKJ4\_up and shJMJD3\_up,  $2.7 \times 10^{-33}$ , and GSKJ4\_down and shJMJD3\_down versus GSKJ4\_up and shJMJD3\_up,  $4.4 \times 10^{-15}$ . \*\*\*, significant.

of primary human samples of paediatric T-ALL using single nucleotide polymorphism (SNP) arrays identified two patients with focal deletions of the *UTX* locus (Fig. 3f). Further targeted sequencing in paediatric and adult T-ALL led to the identification of six more patient cases with *UTX* mutations (Fig. 3g, Extended Data Fig. 5j, k and Supplementary Table 2), including in-frame deletions, missense (Ile598Val) mutations and frameshift alterations. Analysis of bone marrow remission genomic

DNA confirmed the somatic origin of the *UTX* splice site mutation (Extended Data Fig. 5k). Seven out of the eight alterations belonged to male patients, further underlining that the roles of *UTX* and *UTY* do not seem to be interchangeable. These genetic alterations are predicted to have an inactivating role<sup>20,21</sup> and provide further evidence that *UTX* is a tumour suppressor in T-ALL. Indeed, overexpression of *UTX* using a doxycycline-inducible lentiviral system in T-ALL cell lines (Extended Data Fig. 5l) led to suppression of tumour growth and a significant increase in apoptosis (Extended Data Fig. 5m, n).

*Jmjd3*<sup>-/-</sup> mice<sup>29</sup> lack the catalytic domain of the JMJD3 protein (Extended Data Fig. 6a, b) and die perinatally<sup>29</sup>. Haematopoiesis and T-cell development were largely unaffected by the absence of JMJD3 (Extended Data Fig. 6c–h). Genetic ablation of *Jmjd3* in T-ALL led to fewer leukaemic blasts in the peripheral blood, significantly reduced leukaemic cell infiltration of the spleen and liver and improved survival rates in the recipients (Extended Data Fig. 7a–f), consistent with *Jmjd3* having an oncogenic role. These striking phenotypes supported our previous *in vitro* and *in vivo* findings and led us to further explore the therapeutic potential of targeting *JMJD3* activity in T-ALL.

We next tested whether the small molecule GSKJ4 (ref. 5), which is directed against JMJD3 and *UTX* (half-maximum inhibitory concentration (IC<sub>50</sub>) as determined by matrix-assisted laser desorption mass spectroscopy, JMJD3, 18 μM; *UTX* 56 μM; ref. 5), affects maintenance of the disease. We used GSKJ4 at the IC<sub>50</sub> determined for T-ALL cells (2 μM) (Fig. 4a) to treat a panel of T-ALL cell lines. GSKJ4 significantly affected the growth of human T-ALL cell lines and primary human T-ALL cells (T-ALL1–3), leading to cell cycle arrest and increased apoptosis compared with control-inhibitor-treated cells (Fig. 4b and Extended Data Fig. 8a–h). The first detectable changes started at 24 h, and we observed significantly altered phenotypes at 48 h and 72 h (Extended Data Fig. 8i). These GSKJ4 effects appear to be connected to the demethylase activity of JMJD3, as overexpression of catalytically inactive JMJD3 did not rescue the phenotype (Extended Data Fig. 8j, k). The growth of myeloid leukaemia cells, stromal cells and haematopoietic progenitor cells (Extended Data Fig. 8l, m) was unaffected by GSKJ4, demonstrating specificity of function. Mechanistically, we detected gene expression changes starting at 24 h post-GSKJ4 treatment, and significant changes were noted at 48 h and 72 h (Extended Data Fig. 8n) and were coupled to an increase in the H3K27me3 levels at repressed genes (Extended Data Fig. 9a–c). The NOTCH1 and JMJD3 occupancy at specific NOTCH1 target genes that were tested, as well as the total cellular levels of NOTCH1 and JMJD3 and the chromatin H3K27me3 levels, did not significantly change over the treatment duration (Extended Data Fig. 9a–e).

Genome-wide studies identified 486 downregulated genes after 72 h of treatment of human T-ALL cells (CUTLL1) with GSKJ4 (Fig. 4c). There was a significant overlap between the shJMJD3 and GSKJ4 signatures for both downregulated genes ( $P = 4.88 \times 10^{-44}$ ; Fig. 4d and Supplementary Table 3) and upregulated genes ( $P = 2.57 \times 10^{-20}$ ). By contrast, the shUTX-upregulated gene signature significantly overlapped with the GSKJ4-downregulated gene signature. Furthermore, there was a significant overlap between genes upregulated in *Utx* knockout blasts and downregulated by GSKJ4 treatment ( $P = 2.49 \times 10^{-9}$ ; Figs 3e and 4d and Supplementary Table 3), suggesting again that *UTX* and JMJD3 play opposing roles in T-ALL. Genome-wide study of H3K27me3 localization demonstrated that the GSKJ4-downregulated genes experienced gain of H3K27me3 upon GSKJ4 treatment and were marked by the presence of H3K4me3, NOTCH1 and JMJD3 at their promoters (Fig. 4c and Extended Data Fig. 1j). Well-characterized NOTCH1 and JMJD3 targets are highlighted as representative examples of the GSKJ4-downregulated/shJMJD3-downregulated signature and show a significant gain in H3K27me3 upon GSKJ4 treatment (Fig. 4e and Extended Data Fig. 9f). *UTX* was not involved in the regulation of the oncogenic NOTCH1 targets, as revealed by ChIP studies (Extended Data Fig. 9g).

We propose targeting JMJD3 as a novel therapeutic option for paediatric and adult T-ALL. This proposal is based on recent studies<sup>2,4,6</sup> that demonstrate that H3K27me3 catalysed by the PRC2 complex plays

a key role in T-ALL, through antagonism with oncogenic NOTCH1. We demonstrate here that NOTCH1-mediated recruitment of JMJD3 to promoters can explain this antagonism (Extended Data Fig. 10 and see also Supplementary Discussion for extended discussion). We propose that NOTCH1 recruitment leads to PRC2 eviction as a result of the active demethylation of H3K27 through the catalytic activity of JMJD3 and the recruitment of JMJD3 to target promoters. By contrast, the reported increases in the levels of the activating H3K4me3 mark on a large fraction of NOTCH1 target genes<sup>27</sup> (Fig. 4c) can be explained by the fact that NOTCH1 has the ability to participate in MLL complexes (Extended Data Figs 2 and 10). Moreover, we demonstrate the anti-tumorigenic activities of the inhibitor GSKJ4 (ref. 5) and its specificity for T-ALL cells. Clearly, we cannot exclude the possibility that GSKJ4 affects other important epigenetic modulators or signalling pathways. Nevertheless, we consider that the main action of this inhibitor in T-ALL is channelled through the inhibition of JMJD3 activity and propose that such compounds should be tested either as single drugs or in combination with standard chemotherapy.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 25 August 2013; accepted 18 June 2014.**

**Published online 17 August; corrected online 22 October 2014 (see full-text HTML version for details).**

- Baylin, S. B. & Jones, P. A. A decade of exploring the cancer epigenome: biological and translational implications. *Nature Rev. Cancer* **11**, 726–734 (2011).
- Zhang, J. *et al.* The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. *Nature* **481**, 157–163 (2012).
- Jankowska, A. M. *et al.* Mutational spectrum analysis of chronic myelomonocytic leukemia includes genes associated with epigenetic regulation: *UTX*, *EZH2*, and *DNMT3A*. *Blood* **118**, 3932–3941 (2011).
- Ntziachristos, P. *et al.* Genetic inactivation of the polycomb repressive complex 2 in T cell acute lymphoblastic leukemia. *Nature Med.* **18**, 298–303 (2012).
- Kruidenier, L. *et al.* A selective jumoni H3K27 demethylase inhibitor modulates the proinflammatory macrophage response. *Nature* **488**, 404–408 (2012).
- Simon, C. *et al.* A key role for *EZH2* and associated genes in mouse and human adult T-cell acute leukemia. *Genes Dev.* **26**, 651–656 (2012).
- Hubner, M. R. & Spector, D. L. Role of H3K27 demethylases *Jmjd3* and *UTX* in transcriptional regulation. *Cold Spring Harb. Symp. Quant. Biol.* **75**, 43–49 (2010).
- Kooistra, S. M. & Helin, K. Molecular mechanisms and potential functions of histone demethylases. *Nature Rev. Mol. Cell Biol.* **13**, 297–311 (2012).
- Morales Torres, C., Laugesen, A. & Helin, K. *Utx* is required for proper induction of ectoderm and mesoderm during differentiation of embryonic stem cells. *PLoS ONE* **8**, e60020 (2013).
- Wang, C. *et al.* *UTX* regulates mesoderm differentiation of embryonic stem cells independent of H3K27 demethylase activity. *Proc. Natl Acad. Sci. USA* **109**, 15324–15329 (2012).
- Mansour, A. A. *et al.* The H3K27 demethylase *Utx* regulates somatic and germ cell epigenetic reprogramming. *Nature* **488**, 409–413 (2012).
- De Santa, F. *et al.* The histone H3 lysine-27 demethylase *Jmjd3* links inflammation to inhibition of polycomb-mediated gene silencing. *Cell* **130**, 1083–1094 (2007).
- Agger, K. *et al.* The H3K27me3 demethylase *JMJD3* contributes to the activation of the *INK4A-ARF* locus in response to oncogene- and stress-induced senescence. *Genes Dev.* **23**, 1171–1176 (2009).
- Barradas, M. *et al.* Histone demethylase *JMJD3* contributes to epigenetic control of *INK4a/ARF* by oncogenic RAS. *Genes Dev.* **23**, 1177–1182 (2009).
- Jepsen, K. *et al.* SMRT-mediated repression of an H3K27 demethylase in progression from neural stem cell to neuron. *Nature* **450**, 415–419 (2007).
- Sen, G. L., Webster, D. E., Barragan, D. I., Chang, H. Y. & Khavari, P. A. Control of differentiation in a self-renewing mammalian tissue by the histone demethylase *JMJD3*. *Genes Dev.* **22**, 1865–1870 (2008).
- Zhao, W. *et al.* *Jmjd3* inhibits reprogramming by upregulating expression of *INK4a/Arf* and targeting PHF20 for ubiquitination. *Cell* **152**, 1037–1050 (2013).
- Wang, J. K. *et al.* The histone demethylase *UTX* enables RB-dependent cell fate control. *Genes Dev.* **24**, 327–332 (2010).
- Thieme, S. *et al.* The histone demethylase *UTX* regulates stem cell migration and hematopoiesis. *Blood* **121**, 2462–2473 (2013).
- van Haften, G. *et al.* Somatic mutations of the histone H3K27 demethylase gene *UTX* in human cancer. *Nature Genet.* **41**, 521–523 (2009).
- Mar, B. G. *et al.* Sequencing histone-modifying enzymes identifies *UTX* mutations in acute lymphoblastic leukemia. *Leukemia* **26**, 1881–1883 (2012).
- Weng, A. P. *et al.* Activating mutations of *NOTCH1* in human T cell acute lymphoblastic leukemia. *Science* **306**, 269–271 (2004).
- Espinosa, L. *et al.* The Notch/Hes1 pathway sustains NF- $\kappa$ B activation through *CYLD* repression in T cell leukemia. *Cancer Cell* **18**, 268–281 (2010).
- Dik, W. A. *et al.* New insights on human T cell development by quantitative T cell receptor gene rearrangement studies and gene expression profiling. *J. Exp. Med.* **201**, 1715–1723 (2005).
- Van Vlierberghe, P. *et al.* *ETV6* mutations in early immature human T cell leukemias. *J. Exp. Med.* **208**, 2571–2579 (2011).
- Valk, P. J. *et al.* Prognostically useful gene-expression profiles in acute myeloid leukemia. *N. Engl. J. Med.* **350**, 1617–1628 (2004).
- Wang, H. *et al.* Genome-wide analysis reveals conserved and divergent features of Notch1/RBPJ binding in human and murine T-lymphoblastic leukemia cells. *Proc. Natl Acad. Sci. USA* **108**, 14908–14913 (2011).
- Welstead, G. G. *et al.* X-linked H3K27me3 demethylase *Utx* is required for embryonic development in a sex-specific manner. *Proc. Natl Acad. Sci. USA* **109**, 13004–13009 (2012).
- Satoh, T. *et al.* The *Jmjd3-Irf4* axis regulates M2 macrophage polarization and host responses against helminth infection. *Nature Immunol.* **11**, 936–944 (2010).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank the members of the Aifantis laboratory and J. Siegle for discussions throughout the duration of the project; S. Shen for discussions on the analysis of sequencing data; GlaxoSmithKline for the GSKJ4 and GSKJ5 inhibitory compounds; A. Heguy and the NYU Genome Technology Center (supported in part by National Institutes of Health (NIH)/National Cancer Institute (NCI) grant P30 CA016087-30) for assistance with sequencing experiments; the NYU Flow Cytometry facility (supported in part by NIH/NCI grant 5P30CA16087-31) for cell sorting; the NYU Histology Core (5P30CA16087-31) and the NYU mouse facility (NYU Cancer Institute Center Grant 5P30CA16087-31); G. Natoli for providing the anti-JMJD3 antibody; J. Zhang for help with the analysis of the mutation data; and I. Rigo for technical support. I.A. was supported by the NIH (grants 1R01CA133379, 1R01CA105129, 1R01CA149655, 5R01CA173636 and 5R01CA169784). J.N. was supported by the Damon Runyon Cancer Research Foundation. B.K. was supported by the NYU Cell and Molecular Biology Training Program. P.N. was supported by fellowships from Lady Tata Memorial Trust for leukaemia and the American Society of Hematology and NIH/NCI (K99CA188293). T.T. is supported by the NIH training grant 5 T32 CA009161-37. P.V.V. was supported by the Research Foundation Flanders and an Odysseus type II grant. Moreover, this study was supported by an NIH grant (R37-HD04502) to R.J., the ECOG tumour bank, an NIH grant (R01CA120196) to A.A.F.; NCI grants (U24 CA114737 and U10 CA21115) to E.P. and the Stand Up To Cancer Innovative Research Award (A.A.F.). I.A. was also supported by the William Lawrence and Blanche Hughes Foundation, The Leukemia & Lymphoma Society, the Ralph S. French Charitable Foundation Trust, The Chemotherapy Foundation, The V Foundation for Cancer Research and the St. Baldrick's Foundation. I.A. is a Howard Hughes Medical Institute Early Career Scientist. A.T. carried out part of this work while at the Computational Biology Center, IBM Research, Yorktown Heights, New York.

**Author Contributions** I.A. and P.N. designed the experiments and wrote the manuscript. P.N. performed most of the experiments. A.T. designed and performed the analysis of genome-wide data and wrote the manuscript. T.T., E.L., A.S., J.M., B.K., S.B. and J.N. performed experiments and contributed ideas. G.T. provided materials and tips related to the study. A.A.F. and L.X. designed and performed xenograft luciferase experiments and helped with ideas and concepts. P.V.V., E.P., M.S.T., J.M.R. and A.A.F. performed and analysed the mutational studies in adult T-ALL. G.G.W., R.J., T.S. and S.A. provided mouse tissues and helped with ideas and concepts. L.K. and R.P. helped with guidance on the biology and use of GSKJ inhibitors and with manuscript preparation. L.H., J.B. and C.G.M. performed and analysed the mutational studies in paediatric T-ALL.

**Author Information** The high-throughput sequencing data have been deposited in the Gene Expression Omnibus with accession number GSE56696. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to I.A. (iannis.aifantis@nyumc.org), A.T. (Aristotelis.Tsirigos@nyumc.org) or C.G. (charles.mullighan@stjude.org).

## METHODS

**Mice, cell culture and primary cell samples.** The *Jmjd3* (ref. 29) and *Utx*<sup>28</sup> knockout mouse models, as well as the corresponding genotyping strategy, have been described previously. All animals used in this study were treated according to IACUC protocols for the laboratories of I.A., A.A.F. and R.J. The human T-ALL cell lines CUTLL1 (ref. 30), P12-Ichikawa, Loucy, DND41, CEM and Jurkat and the myeloid leukaemia cell lines (THP-1 and HL-60), as well as the mouse T-ALL line (720)<sup>31</sup>, were cultured in RPMI 1640 medium supplemented with 20% FBS and penicillin and streptomycin. All cell lines were tested for the presence of mycoplasma, and only mycoplasma-free lines were used for these studies. Primary human samples were collected by collaborating institutions with informed consent and were analysed under the supervision of the Columbia University Medical Center and St. Jude Children's Research Hospital Institutional Review Boards. The primary cells treated with GSKJ4 inhibitor (for more information on these cells, see ref. 32) were cultured in MEM $\alpha$  medium plus 10% FBS (StemCell Technologies, #06400), 10% human AB<sup>+</sup> serum (Invitrogen), 1% penicillin/streptomycin, 1% GlutaMAX, human interleukin-7 (IL-7) (R&D Systems; 10 ng ml<sup>-1</sup>), human Flt3 ligand (Peprotech; 20 ng ml<sup>-1</sup>), human SCF (Peprotech; 50 ng ml<sup>-1</sup>) and insulin (Sigma; 20 nmol l<sup>-1</sup>). Irradiated MS5 stromal cells overexpressing delta-like 1 (DLL1) were used as a feeder layer, as previously described<sup>33</sup>.

**In vitro drug treatment and shRNA treatment and cell growth, apoptosis and cell cycle analysis.** T-ALL cells were infected twice with shRNA-expressing retroviruses and selected using puromycin. Expression studies took place at different time points during the selection period, and we present the results from day 4 during selection. To calculate the IC<sub>50</sub> of GSKJ4 (GlaxoSmithKline)<sup>5</sup> normalized to the control inhibitor GSKJ5 (GlaxoSmithKline), T-ALL lines were treated with different concentrations of the drug for 5 days. For cell growth, cell lines and primary cultures were treated with 2  $\mu$ M GSKJ4 and GSKJ5 for various times (24 h to 72 h) and stained with annexin V and subjected to cell cycle analysis.  $\gamma$ -Secretase inhibitor ( $\gamma$ SI, specifically Compound E (Alexis Biochemicals)) was used at 500 nM for various periods. For the cell cycle analysis, 5-bromodeoxyuridine (BrdU; 10  $\mu$ M) was added for a 1 h pulse, and incorporation into DNA was determined by using the BrdU Flow Kit (BD Biosciences). Apoptosis was studied by quantification of annexin V staining using the BD Biosciences kit and flow cytometry according to standard protocols provided by the manufacturer. Doxycycline was used at 1  $\mu$ g ml<sup>-1</sup> final concentration.

**Intravenous and subcutaneous xenograft studies.** Studies were conducted as previously published<sup>4</sup>. In both cases, CUTLL1, P12 or CEM T-ALL cells expressing luciferase (FUV-LUC) and the corresponding shRNA (shJMJD3, shUTX or shRenilla) were used. For the intravenous studies, 1  $\times$  10<sup>6</sup> cells were injected retro-orbitally into sublethally irradiated female NRG (NOD Rag1<sup>-/-</sup>Il2rg<sup>-/-</sup>) mice. For subcutaneous studies, 1  $\times$  10<sup>6</sup> cells were mixed with an equal volume of BD Matrigel basement membrane and injected into the flanks of female NOD-SCID mice. In both cases, cell growth was monitored every 2 days using IVIS (Caliper, PerkinElmer).

**Transplantation for reconstitution of the haematopoietic system and for disease progression analysis.** Fetal livers from *Jmjd3*<sup>+/+</sup>, *Jmjd3*<sup>+/-</sup> and *Jmjd3*<sup>-/-</sup> embryos (E13.5, Ly45.2 background) were provided by S.A.'s laboratory, and 1  $\times$  10<sup>6</sup> total (unfractionated) fetal liver cells were used for the reconstitution of the haematopoietic system of lethally irradiated recipients on a Ly45.1 background. Bone marrow was isolated from the recipients, followed by isolation of cells of the Ly45.2 background using flow cytometry. Total Ly45.2 bone marrow mononuclear cells (2.5  $\times$  10<sup>5</sup> cells) were mixed with equal numbers of Ly45.1 (wild-type) bone marrow cells and transplanted into lethally irradiated recipients to study haematopoietic reconstitution in a competitive setting.

For the *Utx*<sup>+/+</sup>, *Utx*<sup>+/-</sup> and *Utx*<sup>-/-</sup> (Ly45.2) background, 2.5  $\times$  10<sup>5</sup> cells of total Ly45.2 bone marrow mononuclear cells were mixed with equal numbers of Ly45.1 (wild-type) bone marrow cells and transplanted into lethally irradiated recipients to study haematopoietic reconstitution in a competitive setting similar to the *Jmjd3* study.

In both cases, reconstitution of the haematopoietic system was monitored by analysis of the peripheral blood for the main haematopoietic lineages. The thymus and spleen of some recipients were isolated and analysed at 3 months post transplantation.

For analysis of leukaemia progression, c-Kit<sup>+</sup> haematopoietic progenitors from the bone marrow of both *Jmjd3* and *Utx* knockout models were magnetically selected (STEMCELL Technologies) using an antibody against CD117 (c-Kit) and were cultured overnight in the presence of 50 ng ml<sup>-1</sup> SCF, 50 ng ml<sup>-1</sup> Flt3 ligand, 10 ng ml<sup>-1</sup> IL-3 and 10 ng ml<sup>-1</sup> IL-6. Overexpression of oncogenic *Notch1* mutants (the intracellular part of NOTCH1 (NOTCH1-IC) and DeltaE (NOTCH1- $\Delta$ E)) in bone marrow haematopoietic progenitors followed by transplantation into mouse recipients led to the development of T-ALL, characterized by the presence of leukaemic blasts in the peripheral blood that infiltrated the peripheral lymphoid organs, progressively leading to the death of the animals (Extended Data Fig. 5c). The cells

were infected with NOTCH1-IC or NOTCH1- $\Delta$ E (and green fluorescent protein (GFP)) expressing retroviruses twice (24 h and 48 h post c-Kit selection). Viral transduction efficiency was determined by measuring reporter fluorescence over a total period of 4 days, and total populations were transferred via retro-orbital injection into lethally irradiated congenic recipients along with 2.5  $\times$  10<sup>5</sup> total (wild-type) bone marrow mononuclear cells for haemogenic support. GFP<sup>+</sup> cells (4  $\times$  10<sup>4</sup>) were transplanted in both NOTCH1-IC and NOTCH1- $\Delta$ E studies. The Mantel-Cox test was used for the analysis of the survival data. No randomization or blinding method was used during these animal studies.

**Antibodies, reagents, kits and virus production.** Protein-G-coated magnetic beads were purchased from Invitrogen. Antibodies against the following proteins were used: monoclonal mouse H3K27me3 (histone H3 migrates at around 17 kDa) (Abcam, ab6002), monoclonal mouse H3K27me1 (Active Motif, 61015), polyclonal rabbit H3K4me3 (Active Motif, 39159), polyclonal rabbit NOTCH1 (the intracellular part of the protein migrates at around 110 kDa), polyclonal rabbit JMJD3 (protein migrates at around 170 kDa) (Abgent, AP1022a (human) and AP1022b (mouse)), as well as polyclonal rabbit JMJD3 (Cell Signaling Technology, 3457), polyclonal rabbit UTX (protein migrates at around 160 kDa) (Abcam, ab36938, and Bethyl Laboratories, A302-374A), polyclonal rabbit NF- $\kappa$ B (p65, protein migrates at around 65 kDa) (Santa Cruz Biotechnology, sc-109 and sc-372) and control IgG (Santa Cruz Biotechnology, msc-2025 (mouse) and sc-2027 (rabbit)). All antibodies for flow cytometry were from eBioscience. All antibodies used had been tested and shown to be specific for the purposes we used them for by the suppliers. The acid extraction protocol by Abcam was used for the characterization of histone mark levels upon GSKJ4 treatment. To generate the virus, we infected 293T cells with a plasmid expressing the shRNA (an miR-30-based system<sup>34</sup>) against *JMJD3* or *UTX* (shJMJD3A, 5'-CAGGGAAGTTTCGAGAAGTCCTATAGTGAAGCCACAGATGTATAGGACTCTCGAAGTCCCTT-3'; shJMJD3B, 5'-ACACCAGCAGTAGCAACAGCAATAGTGAAGCCACAGATGTATTGCTGTGCTACTGCTGGTGG-3'; shUTX, 5'-ACACAAGGTAGTCTACAGAATATAGTGAAGCCACAGATGTATATTCGTAGACTACCTTGTGG-3'). We also used shRenilla (5'-CTCGAGAAGGTAATTGCTGTGTGACAGTGAGCGCAGGAATTATAATGCTTATAGTGAAGCCACAGATGTATAGTATAGTAAGCATTATTCCTAGCTCTCGAATTC-3') as a control and the retroviral packaging plasmid. Viral supernatant was collected over a period of 72 h and used for the transduction of T-ALL cells. The cells were infected twice and then selected with puromycin starting 3 days after viral infection. Reporter fluorescence was used (as determined by flow cytometry) for the quantification of shRNA.

**Histopathology.** Organs were harvested from the animals and immersion fixed with 4% paraformaldehyde<sup>4</sup> overnight at 4 °C. Samples were washed with PBS three times for 1 h at room temperature and dehydrated in 70% ethanol. Samples were embedded in paraffin blocks. Sections (6- $\mu$ m thick) were stained with haematoxylin and eosin following standard procedures. Peripheral blood smears were briefly fixed in methanol and stained with Wright-Giemsa solution (Fisher). Slides were rinsed with water, dried, mounted with Cytoseal 60 and coverslipped.

**Protein immunoprecipitation for interaction studies.** For the interaction studies between the NOTCH1 complex (NOTCH1 and MAML1) and the epigenetic modulators (UTX, JMJD3 and EZH2), we used standard protocols used elsewhere.

In brief, cells were resuspended in TENT buffer (50 mM Tris, pH 8.0, 5 mM EDTA, 150 mM NaCl and 0.05% (v/v) Tween-20) supplemented with the inhibitors at a concentration of 20  $\times$  10<sup>6</sup> cells ml<sup>-1</sup> buffer. Cell lysates were passed through a 25G syringe five times and incubated on ice for 30 min, followed by centrifugation to remove cell debris (5 min, 13,000g). The cleared lysate was precleared with beads for 1 h at 4 °C to decrease non-specific binding and incubated overnight with the corresponding antibody-bound bead complexes. Five micrograms of antibody was used for 3 mg of extracts.

**RNA-seq library preparation and analysis.** Whole RNA was extracted from 1–5  $\times$  10<sup>6</sup> T-ALL cells or primary cells using the RNeasy kit (QIAGEN) according to the manufacturer's protocol. Poly(A)<sup>+</sup> RNA was enriched using magnetic oligo(dT)-containing beads (Invitrogen). cDNA was prepared and strand-specific libraries were constructed using the dUTP method as described previously<sup>35</sup>. Libraries were sequenced on the Illumina HiSeq 2000 using the 50-base-pair single-read method.

**ChIP and ChIP-seq library preparation.** ChIP experiments were performed as described previously<sup>4</sup>. In brief, for the analysis of histone marks, we fixed the cells with 1% formaldehyde for 10 min at 25 °C and lysed them by the addition of nucleus incubation buffer (15 mM Tris, pH 7.5, 60 mM KCl, 150 mM NaCl, 15 mM MgCl<sub>2</sub>, 1 mM CaCl<sub>2</sub>, 250 mM sucrose and 0.3% NP-40) and incubation at 4 °C for 10 min. The nuclei were washed once with digest buffer (10 mM NaCl, 10 mM Tris, pH 7.5, 3 mM MgCl<sub>2</sub> and 1 mM CaCl<sub>2</sub>), and we used micrococcal nuclease (USB) in digest buffer to generate mononucleosomal particles. The reaction was stopped by the addition of EDTA (20 mM). The nuclei were lysed in nucleus lysis buffer (50 mM Tris-HCl, pH 8.0, 10 mM EDTA, pH 8.0, and 1% SDS) followed by sonication using a Bioruptor (Diagenode), and chromatin was precleared by the addition of nine



volumes of IP dilution buffer (0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, pH 8.0, 16.7 mM Tris-HCl, pH 8.0, and 167 mM NaCl) and magnetic Dynabeads. One per cent of the chromatin was kept as input. We coupled 2.5 µg antibody with 25 µl of beads for 4 h in reaction buffer, and the complex was added to precleared chromatin (the equivalent of  $10^5$ – $10^6$  cells, depending on the antibody) followed by overnight incubation at 4 °C with rotation. We washed the complexes bound to the beads using buffers with increasing salt concentration: once with wash A (20 mM Tris-HCl, pH 8, 150 mM NaCl, 2 mM EDTA, 1% (w/v) Triton X-100 and 0.1% (w/v) SDS), once with wash B (20 mM Tris-HCl, pH 8.0, 500 mM NaCl, 2 mM EDTA, 1% (w/v) Triton X-100 and 0.1% (w/v) SDS), once with wash C (10 mM Tris-HCl, pH 8.0, 250 mM LiCl, 1 mM EDTA, 1% (w/v) NP-40 and 1% (w/v) deoxycholic acid) and twice with TE, followed by treatment with RNase and proteinase K. The cross-links were then reversed, and the DNA was precipitated using ethanol and glycogen.

For JMJD3 ChIP, the cells were fixed with 1% formaldehyde for 10 min at 25 °C and lysed on ice using 1 ml cell lysis buffer (50 mM HEPES-KOH, pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40 and 0.25% Triton X-100) per  $1 \times 10^7$  cells. We resuspended the pellet in 1 ml buffer II (10 mM Tris-HCl, pH 8, 200 mM NaCl, 1 mM EDTA, pH 8, and 0.5 mM EGTA) per  $1 \times 10^7$  cells. We further resuspended the nuclei in buffer III (10 mM Tris-HCl, pH 8, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% sodium deoxycholate and 0.5% n-lauroylsarcosine) and sonicated the solution with a Bioruptor for 40 min. Triton X-100 was added to a final concentration of 1%, and the chromatin preparation was precleared using magnetic beads. The antibody (5 µg) was coupled to the magnetic beads (50 µl) as in the case of the histone marks, and the complex was added to the precleared chromatin (the equivalent of  $1 \times 10^7$  cells per reaction). The reaction mix was then incubated for 12–16 h. The beads with the immunoprecipitated chromatin fragments were washed eight times with RIPA buffer (50 mM HEPES-KOH, pH 7.6, 300 mM LiCl, 1 mM EDTA, 1% NP-40 (IGEPAL) and 0.7% sodium deoxycholate) and once with TE. The DNA was cleaned as in the case of the chromatin marks (see above). Libraries were generated as described previously<sup>4</sup>, including end repair, A-tailing, adaptor ligation (Illumina TruSeq system) and PCR amplification of the libraries. AMPure XP beads (Beckman Coulter, A63880) were used for DNA cleaning in each step of the process.

**Sequence analysis of primary samples.** Sequencing and analysis of paediatric T-ALL samples was conducted as described in previously published studies<sup>2,36</sup>. In brief, sequencing of *UTX* in the paediatric T-ALL cohort was performed by PCR of whole genome amplified DNA, followed by sequencing using 3730xl instruments (Applied Biosystems) as previously described<sup>37</sup>. Single nucleotide variations were detected by SNPdetector<sup>38</sup> and PolyScan<sup>39</sup> and validated by sequencing of both tumour and matched non-tumour samples. A total of 107 paediatric patients were screened, including 64 cases with ETP ALL (25 females and 39 males) and 43 with 'typical' T-ALL (8 females and 35 males). *UTX* mutations were detected in 4.7% of the total population and in 6.8% of the male population. No *UTX* mutations were detected in female samples. The two deletions and one of the frameshift mutations were found in patients with typical T-ALL, and the other two in patients with ETP ALL.

Regarding the adult T-ALL case, all 83 samples were collected in the Eastern Cooperative Oncology Group (ECOG) clinical trials E2993 (ref. 40) and C10403 and analysed under the supervision of the Columbia University Medical Center Institutional Review Board. Informed consent to use leftover material for research purposes was obtained from all of the patients at trial entry in accordance with the Declaration of Helsinki. All exon sequences from *UTX* were amplified from genomic DNA by PCR and analysed by direct dideoxynucleotide sequencing. The primer sequences used for *UTX* sequencing have been described previously<sup>20</sup>.

**Data sources and computational tools.** Patient and physiological T-cell expression data were obtained from refs 2, 24, 41. Human genome assembly version hg19/GRCh37 and Ensembl annotation release 69 were used for the RNA-seq, ChIP-seq and data integration analyses. NOTCH1, RBP-J, H3K4me3 and H3K27me3 ChIP-seq data for CUTLL1 cells were obtained from ref. 27. For the functional enrichment analysis, MSigDB<sup>42</sup> version 3.1 was used. Bowtie<sup>43</sup> version 0.12.7 was used for alignment of sequenced reads. RNA-seq data analysis was performed using DEGseq<sup>44</sup>. MACS<sup>45</sup> version 2.0.10 was used for JMJD3 ChIP-seq peak discovery, in conjunction with the irreproducible discovery rate (IDR) method<sup>46</sup>. GenomicTools<sup>47</sup> version 2.7.2 was used for performing genomic interval mathematical operations, genomic interval annotation, H3K27me3 ChIP-seq comparisons (GSKJ4 versus control) and ChIP-seq heatmap generation.

**Expression analysis of primary samples.** Processed T-ALL and B-ALL patient microarray expression data were downloaded from ref. 2 (GEO accession GSE33315), physiological T-cell expression data from ref. 24 (GEO accession GSE22601) and acute myeloid leukaemia (AML) expression data from ref. 41 (GEO accession GSE6891). Data were first converted to the logarithmic scale when necessary and then quantile-normalized across samples. The Wilcoxon two-sided unpaired test per gene probe was used to determine significant differences between sample categories (T-ALL, B-ALL, AML and physiological T cells; Fig. 1e). A gene was considered

significantly overexpressed in T-ALL compared with the rest of the sample categories if at least one of its associated probes was significantly overexpressed in T-ALL according to the statistical test.

Genes experiencing loss of H3K27me3 at TSSs in our mouse NOTCH-IC model compared with normal double-positive (DP) mouse cells were obtained from our previous study<sup>4</sup>. Enrichment of human homologues of these genes in JMJD3-correlating genes in the patient data described above was estimated as follows. First, Pearson's correlation of JMJD3 expression (separately for each JMJD3 probe) against expression of each gene was computed. Then, the distribution of the correlations of the genes losing H3K27me3 (human homologues of the mouse genes) was compared with that of the genes that did not lose H3K27me3, using Student's *t*-test (separately for each JMJD3 probe, minimum *P* value shown in the corresponding figure (Fig. 1f)) or the Wilcoxon one-sided unpaired test (data not shown), yielding similar results. This analysis was repeated for NF-κB1, NF-κB2, REL, RELA, RELB, HES1, UTX and EZH2 (Extended Data Fig. 1i).

**JMJD3 peak identification, characterization and overlap with published data sets.** JMJD3 ChIP-seq reads were aligned using Bowtie (with default parameters, except for *-m* 1 so as to report only unique alignments) on human assembly version hg19. Peak discovery was performed with MACS (version 2.0.10) using default parameters, except for using a fragment size of 300 base pairs as estimated with the Agilent 2100 Bioanalyzer. Sonicated input was used as a control for peak discovery. Then, we used the IDR method<sup>46</sup>, guidelines and pipeline available for narrow peaks at the URL <https://sites.google.com/site/anshulkundaje/projects/idr> to determine highly reproducible peaks supported by both JMJD3 replicates.

JMJD3 peaks were characterized according to their genome-wide distribution (Extended Data Fig. 1k) into the following groups: (a) 1-kilobase (kb) TSS-flanking regions of transcript isoforms; (b) gene body regions (excluding any regions overlapping with (a)); and (c) upstream regions of a minimum of 10 kb and a maximum of 100 kb (excluding any regions overlapping with (a) or (b)).

Co-occurrence of JMJD3 peaks with H3K4me3, H3K27me3, NOTCH1 and RBP-J was computed as the percentage of such peaks (5,000 top-scoring peaks for each protein obtained from ref. 27; GEO accession GSE29600) that have some overlap with a JMJD3 peak. The statistical significance of these overlaps was determined using random resampling simulation (for example, H3K4me3 peaks were randomly redistributed along the genome). As a control, we used the percentage of TSSs that have JMJD3 peaks (this is a rather conservative control since genome-wide JMJD3 occupancy is much lower, as a result of JMJD3 being concentrated in TSSs), and compared with this control, we obtained an ~7-fold enrichment of H3K4me3–JMJD3 ( $P < 0.001$  as determined by the random resampling scheme). Similar enrichments were obtained for NOTCH1–JMJD3 and RBP-J–JMJD3 co-occurrence, whereas no significant enrichment was observed for H3K27me3-silenced or H3K4me1 enhancer-related regions (Extended Data Fig. 1j).

**RNA-seq analysis.** Differential gene expression analysis was performed for each matched knockdown versus control pair, separately in each biological or technical replicate in each of two cell lines (CUTLL1 and CEM). Three types of comparisons were tested: (1) *JMJD3* knockdown versus Renilla; (2) *JMJD3* knockdown versus *UTX* knockdown; and (3) *UTX* knockdown versus Renilla. DEGseq<sup>44</sup> was used to analyse (a) matched knockdown–Renilla replicates in separate DEGseq runs and (b) all replicates on a combined DEGseq run. For the mouse (*Utx* knockout) samples, spleen and bone marrow from a wild-type male (referred to as animal #9), as well as spleen from a wild-type female (animal #10), were compared with spleen and bone marrow from a knockout male (#23) and spleen from another knockout male (#27) (see also our GEO accession GSE56696). For illustration, scatter plots (Fig. 2d and Extended Data Fig. 3g–i) were created using values obtained from DEGseq analysis of merged biological and/or technical replicates. Gene RNA-seq FPKM values were computed using GenomicTools<sup>47</sup>. The *P*-value cutoff for differential expression was set at  $1 \times 10^{-5}$ , with the minimum absolute log<sub>2</sub> fold change set at 0.5. However, all key results in this study (that is, the significance of the overlaps of the various gene expression signatures demonstrating the contrasting roles of JMJD3 and UTX) are robust to changes in these two parameters (data not shown).

The *P* value of a gene set of size *t* (for example, GSKJ4-downregulated genes) containing *k* genes with a specific attribute (for example, shJMJD3-mediated downregulation or *UTX* knockout upregulation) was determined against the null hypothesis that *k* or more such genes could have been observed merely by chance in an equal sized gene set that was randomly drawn from the entire reference set of genes of size *N* (that is, all downregulated, upregulated and constant genes). This *P* value was obtained by using the hypergeometric cumulative distribution with parameters *N*, *t*, *k* and *n*, where *n* is the number of genes possessing the attribute in the entire reference gene set of size *N*.

**H3K27me3 gain and loss analysis.** JMJD3-affected (upregulated or downregulated) genes were defined as genes whose expression was significantly differentially expressed in *JMJD3* knockdown cells compared with both Renilla and *UTX* knockdown cells. Changes in JMJD3 binding and the H3K27me3 mark around gene



TSSs between cells treated with the inhibitor GSKJ4 and the control GSKJ5 were determined using GenomicTools (“genomic\_apps peakdiff” tool) as described in a previously published study<sup>4</sup>. Epigenetic changes between the treatment (shJMJD3 or GSKJ4) and control samples were determined by evaluating sliding windows across the genome using the following protocol. First, enriched ChIP-seq windows were identified separately for each of the two samples under comparison using a window-based approach and the binomial probability distribution to compare signal reads with control reads in each window. Subsequently, for each genomic window enriched in at least one of the two samples, the total number of reads was determined, and the window read counts were normalized using quantile normalization across biological replicates and samples before comparison. Finally, for each window, the fold change between the two samples was calculated (GSKJ4 versus control, and vice versa). To estimate the false discovery rate, the distribution of the observed H3K27me3 fold changes was compared with the distribution of fold changes between replicates of the same treatment. This comparison was performed independently at different H3K27me3 read density levels to control for artificially high fold changes due to low read counts in the denominator. Significant epigenetic changes are reported at 5% false discovery.

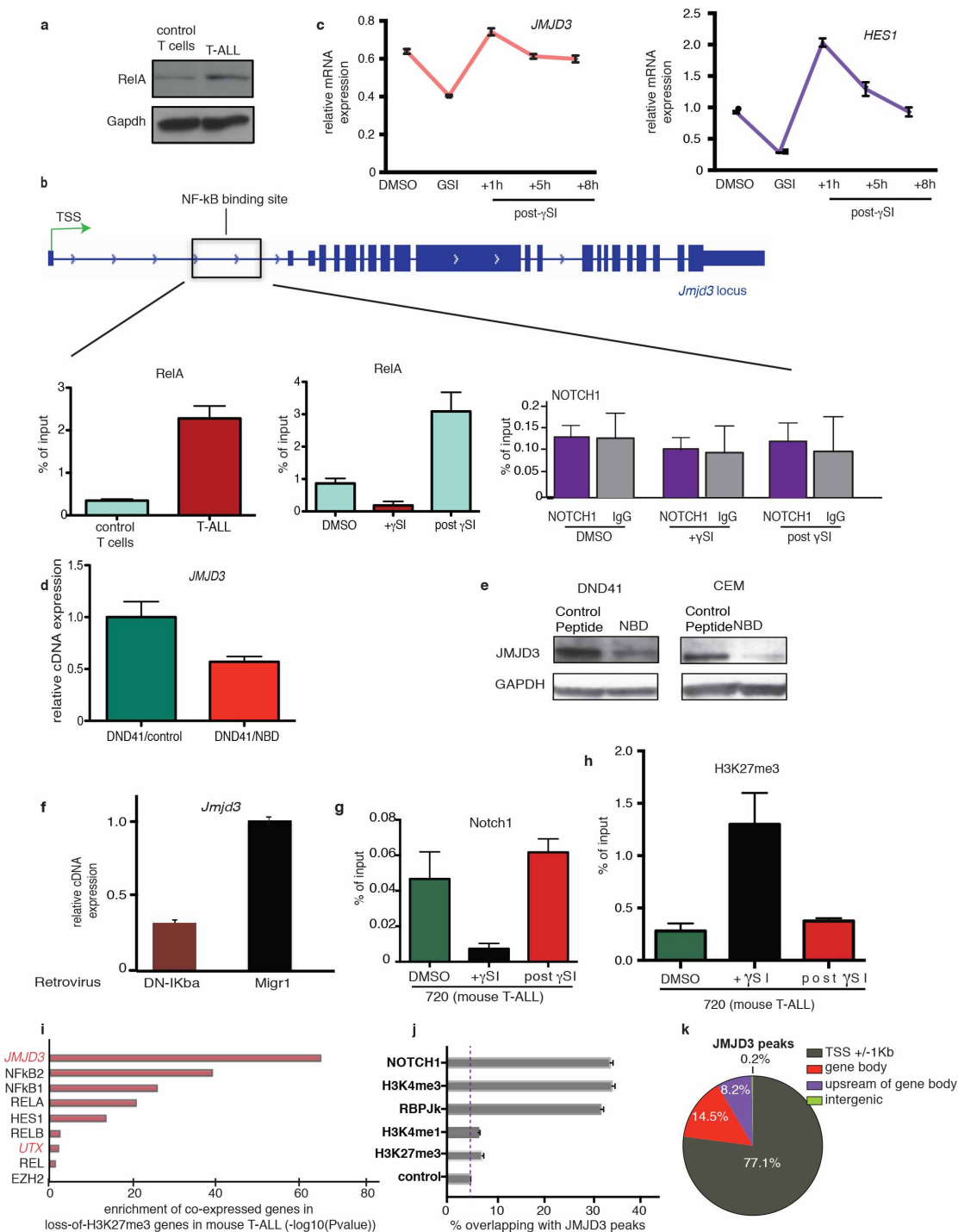
JMJD3, NOTCH1, H3K4me3 and H3K27me3 heatmaps were generated using GenomicTools (“genomic\_apps heatmap” utility) over log-transformed read counts in 200-nucleotide non-overlapping bins of 4-kb-flanked TSSs. Box plots of H3K27me3  $\log_2$ [fold changes] (GSKJ4 versus control) show the distribution of values in (a) JMJD3 targets, (b) commonly downregulated genes upon shJMJD3 and GSKJ4 treatment, and (c) the intersection of GSKJ4-upregulated and shJMJD3-upregulated genes as a negative control. *P* values were computed using a one-sided Wilcoxon unpaired test for (a) and (b) versus the control (c).

**RNA-seq and ChIP-seq replicate reproducibility.** For RNA-seq experiments, we focused on the reproducibility of gene expression levels as measured by FPKM values. For each pair of replicates, we computed the Spearman and Pearson correlations, as well as the Pearson correlation on log-transformed FPKM values. In general, Pearson correlations were much higher because higher values are dominant, and highly expressed genes tend to be more reproducible. Using a Pearson correlation on log-transformed values attempts to balance the expression distribution and allow contributions from genes that are expressed at a lower level, thereby providing a more realistic genome-wide reproducibility metric. Spearman correlations focus on the ranking of gene expression, and in our experiments, in general, were a more conservative (that is, lower) and consistent (lower variability across various settings, and when comparing different cell lines (that is, CUTLL1 and CEM)) estimate of reproducibility; therefore, for simplicity, we have reported only the Spearman correlations.

For ChIP-seq ‘broad peak’ experiments (H3K27me3), we also used Pearson, log-transformed Pearson and Spearman correlations on (a) TSSs and (b) all genome-wide peaks. As before, Spearman correlation was the most conservative and consistent estimate of reproducibility.

For ChIP-seq ‘narrow peak’ experiments (JMJD3), in addition to TSS-based and genome-based correlations, we used the IDR method<sup>46</sup>, guidelines and pipeline available for narrow peaks at the URL <https://sites.google.com/site/anshulkundaje/projects/idr>. Apart from determining the reproducibility, we also used the IDR method to determine high-confidence peaks supported by both JMJD3 replicates.

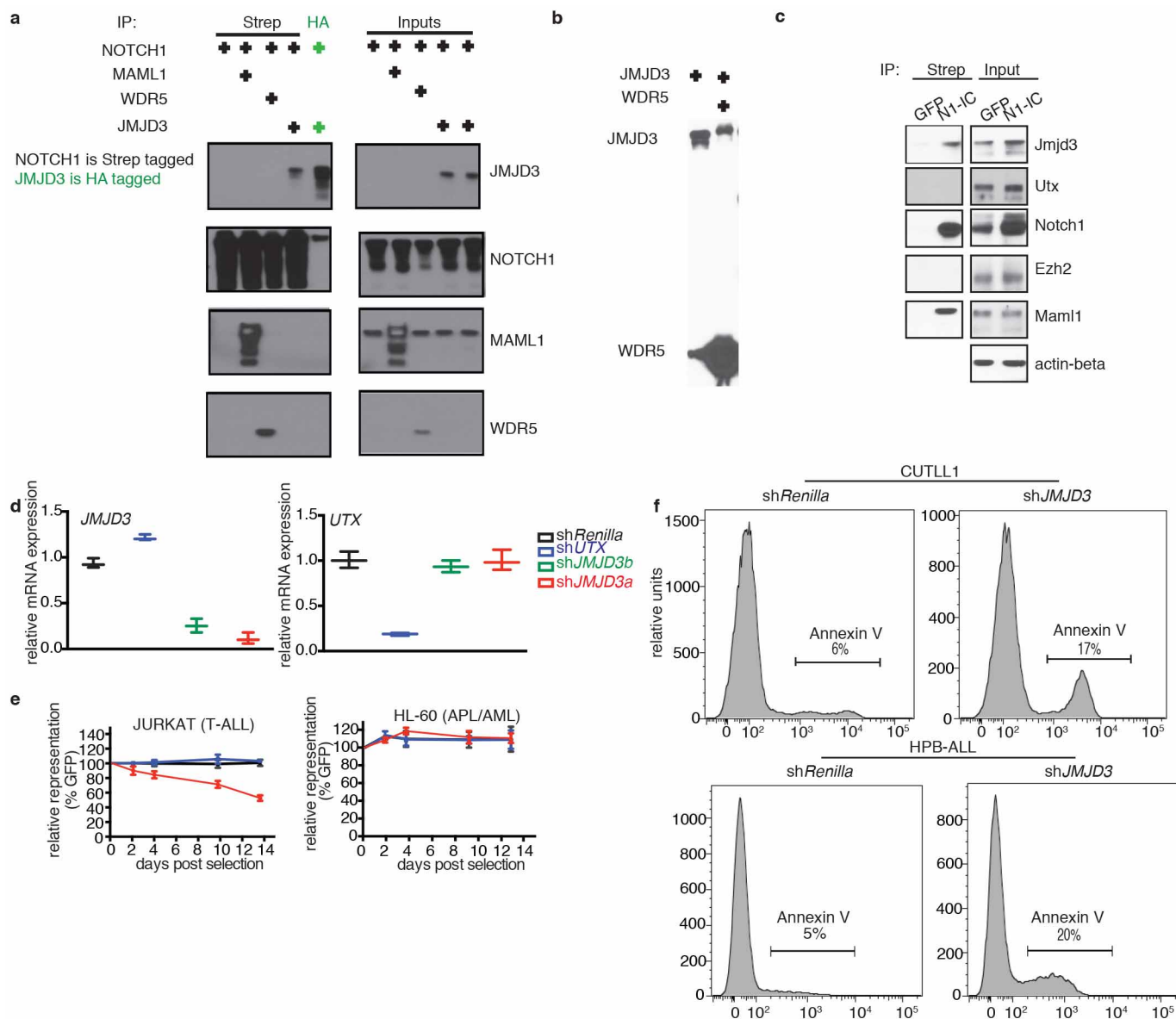
30. Palomero, T. *et al.* CUTLL1, a novel human T-cell lymphoma cell line with t(7;9) rearrangement, aberrant NOTCH1 activation and high sensitivity to  $\gamma$ -secretase inhibitors. *Leukemia* **20**, 1279–1287 (2006).
31. Sharma, V. M. *et al.* Notch1 contributes to mouse T-cell leukemia by directly inducing the expression of *c-myc*. *Mol. Cell. Biol.* **26**, 8022–8031 (2006).
32. King, B. *et al.* The ubiquitin ligase FBXW7 modulates leukemia-initiating cell activity by regulating MYC stability. *Cell* **153**, 1552–1566 (2013).
33. Armstrong, F. *et al.* NOTCH is a key regulator of human T-cell acute leukemia initiating cell activity. *Blood* **113**, 1730–1740 (2009).
34. Dickins, R. A. *et al.* Probing tumor phenotypes using stable and regulated synthetic microRNA precursors. *Nature Genet.* **37**, 1289–1295 (2005).
35. Zhong, S. *et al.* High-throughput Illumina strand-specific RNA sequencing library preparation. *Cold Spring Harb. Protoc.* **2011**, 940–949 (2011).
36. Mullighan, C. G. Single nucleotide polymorphism microarray analysis of genetic alterations in cancer. *Methods Mol. Biol.* **730**, 235–258 (2011).
37. Mullighan, C. G. *et al.* CREBBP mutations in relapsed acute lymphoblastic leukaemia. *Nature* **471**, 235–239 (2011).
38. Zhang, J. *et al.* SNPdetector: a software tool for sensitive and accurate SNP detection. *PLOS Comput. Biol.* **1**, e53 (2005).
39. Chen, K. *et al.* PolyScan: an automatic indel and SNP detection approach to the analysis of human resequencing data. *Genome Res.* **17**, 659–666 (2007).
40. Marks, D. I. *et al.* T-cell acute lymphoblastic leukemia in adults: clinical features, immunophenotype, cytogenetics, and outcome from the large randomized prospective trial (UKALL XII/ECOG 2993). *Blood* **114**, 5136–5145 (2009).
41. Verhaak, R. G. *et al.* Prediction of molecular subtypes in acute myeloid leukemia based on gene expression profiling. *Haematologica* **94**, 131–134 (2009).
42. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
43. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
44. Wang, L., Feng, Z., Wang, X., Wang, X. & Zhang, X. DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* **26**, 136–138 (2010).
45. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
46. Li, Q., Brown, J. B., Huang, H. & Bickel, P. J. Measuring reproducibility of high-throughput experiments. *Ann. Appl. Stat.* **5**, 1752–1779 (2011).
47. Tsigoris, A., Haiminen, N., Bilal, E. & Utro, F. GenomicTools: a computational platform for developing high-throughput analytics in genomics. *Bioinformatics* **28**, 282–283 (2012).



### Extended Data Figure 1 | JMJD3 is induced through activation of the NF- $\kappa$ B pathway in a NOTCH1-dependent mode in T-ALL and binds to NOTCH1 target genes.

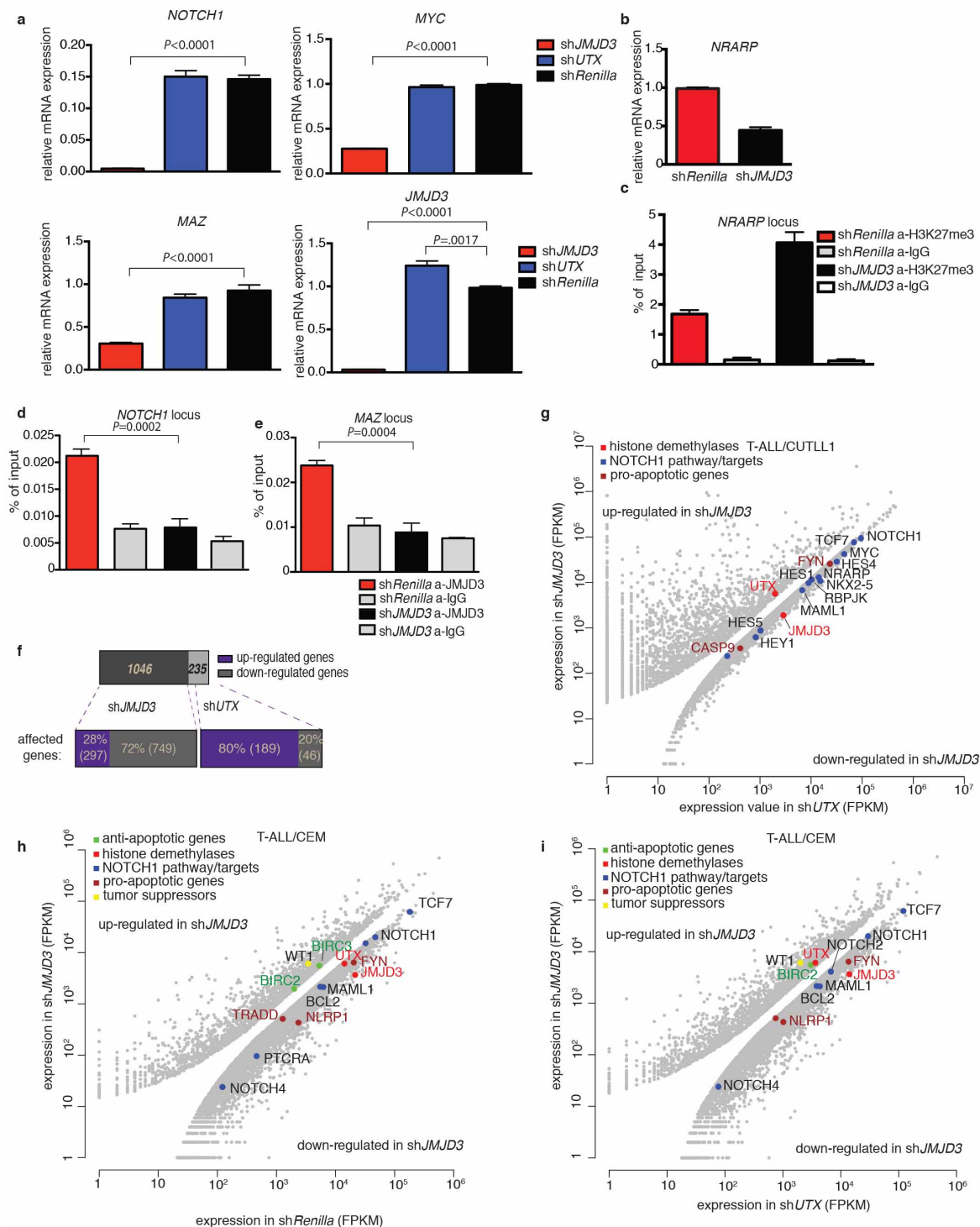
**a**, Levels of p65 (RELA) protein in control T cells and T-ALL tumour cells. A representative sample from three mice is shown. **b**, Schematic representation of the *Jmjd3* locus showing the p65 binding site (upper) and ChIP analysis for p65 binding to the *Jmjd3* locus in mouse control T cells and T-ALL tumour cells, as well as T-ALL cells upon treatment with  $\gamma$ -secretase inhibitor ( $\gamma$ SI), which affects NOTCH1 levels (centre). NOTCH1 binding to this region upon  $\gamma$ SI treatment in T-ALL cells is also shown (right). **c**, Analysis of *JMJD3* and *HES1* messenger RNA levels upon  $\gamma$ SI treatment of CUTLL1 cells. The average of three independent studies is shown. **d, e**, Expression levels of the *JMJD3* transcript (**d**) and protein (**e**) upon treatment of human T-ALL lines (DND41 and CEM) with a NEMO binding

domain (NBD) inhibitor of the NF- $\kappa$ B pathway. **f**, *JMJD3* levels in T-ALL cells upon inhibition of the NF- $\kappa$ B pathway using a dominant negative form of I $\kappa$ B $\alpha$  (DN-I $\kappa$ B $\alpha$ ). **g, h**, ChIP for NOTCH1 (**g**) and H3K27me3 (**h**) on the *Hes1* promoter upon  $\gamma$ SI treatment of mouse T-ALL cells. In **d** and **f-h**, the average of three studies is shown. In **e**, a representative example from three studies is shown. **i**, Genes correlated with selected human genes (including *JMJD3* and *NFKB1*) were tested for enrichment in loss-of-H3K27me3 genes during the transition to T-ALL in the mouse model. **j**, Overlap of *JMJD3* peaks with peaks of important activating (H3K4me3 and H3K4me1) and repressive (H3K27me3) epigenetic marks, as well as members of the NOTCH1 complex. The percentage of TSSs containing *JMJD3* peaks was used as a conservative control and is an alternative to the much lower genome-wide *JMJD3* occupancy. **k**, Genome-wide distribution of *JMJD3* peaks in human T-ALL.



**Extended Data Figure 2 | JMJD3 is vital for T-ALL growth through participation in NOTCH1 transcriptional programs.** **a**, NOTCH1 interaction analyses for JMJD3, MAML1 and WDR5 proteins in 293T cells. Interaction with JMJD3 was confirmed in a reciprocal way (right-most lane, immunoprecipitation (IP) using an anti-haemagglutinin (HA) antibody). **b**, Expression of JMJD3 and WDR5 in 293T cells, followed by immunoprecipitation using the anti-HA antibody against HA-JMJD3. An anti-Flag antibody was used for the detection of both proteins. **c**, NOTCH1 interaction studies for JMJD3 and MAML1 proteins in mouse T-ALL cells expressing a Flag/Strep form of intracellular NOTCH1. StrepTactin beads were used for NOTCH1 precipitation in the absence of detectable intracellular NOTCH1, and different antibodies were used for the detection of JMJD3,

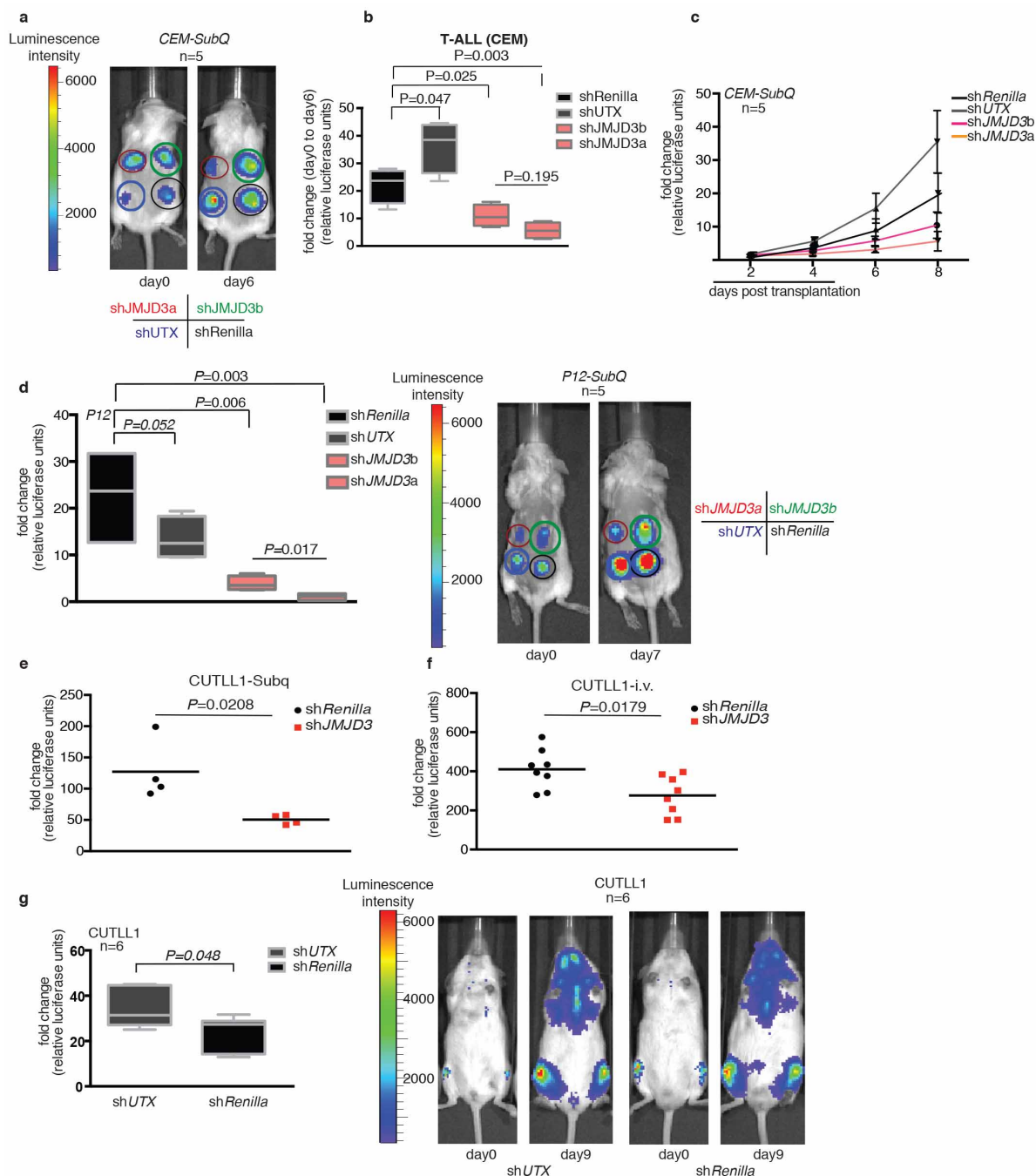
MAML1, EZH2 and UTX. Extracts from green fluorescent protein (GFP)-expressing cells were used as negative control. All experiments were repeated three times (biological replicates), and a representative example is shown. **d**, mRNA expression of *JMJD3* and *UTX* upon treatment with shRNA against *JMJD3* or *UTX*. The expression after treatment of CEM cells with two shRNAs against *JMJD3* and one shRNA against *UTX* and one control (Renilla) is shown. **e**, The effects on cell proliferation as measured by the loss of GFP-expressing shRNA. HL-60 is an acute promyelocytic leukaemia cell line (APL), which is a subtype of acute myeloid leukaemia (AML) and is used as control in this study. For both cell lines, the average results from three representative studies are shown. **f**, Annexin V staining upon shJMJD3 and shRenilla treatment of CUTLL1 cells (top) and HPB-ALL cells (bottom).



**Extended Data Figure 3 | JMJD3 binds to genes with important oncogenic functions and is vital for T-ALL growth.** **a**, *JMJD3* but not *UTX* genetic inactivation impairs the expression of important oncogenic genes. *NOTCH1*, *MYC* and *MAZ*, as well as *JMJD3*, expression levels are shown. shUTX treatment results in significant upregulation of *JMJD3* compared with shRenilla (control)-treated cells. The average results from three studies are shown. **b**, Significant expression changes in *NRARP* transcript levels upon *JMJD3* knockdown. **c**, ChIP for H3K27me3 on the *NRARP* locus. **d**, **e**, Binding of *JMJD3* to the *NOTCH1* (**d**) and *MAZ* (**e**) promoters upon shJMJD3 and shRenilla (control) treatment. The average results from three studies are shown. **f**, Numbers of upregulated and downregulated genes are shown for shJMJD3- and shUTX-treated cells compared with shRenilla-treated cells. **g**, Scatter plot

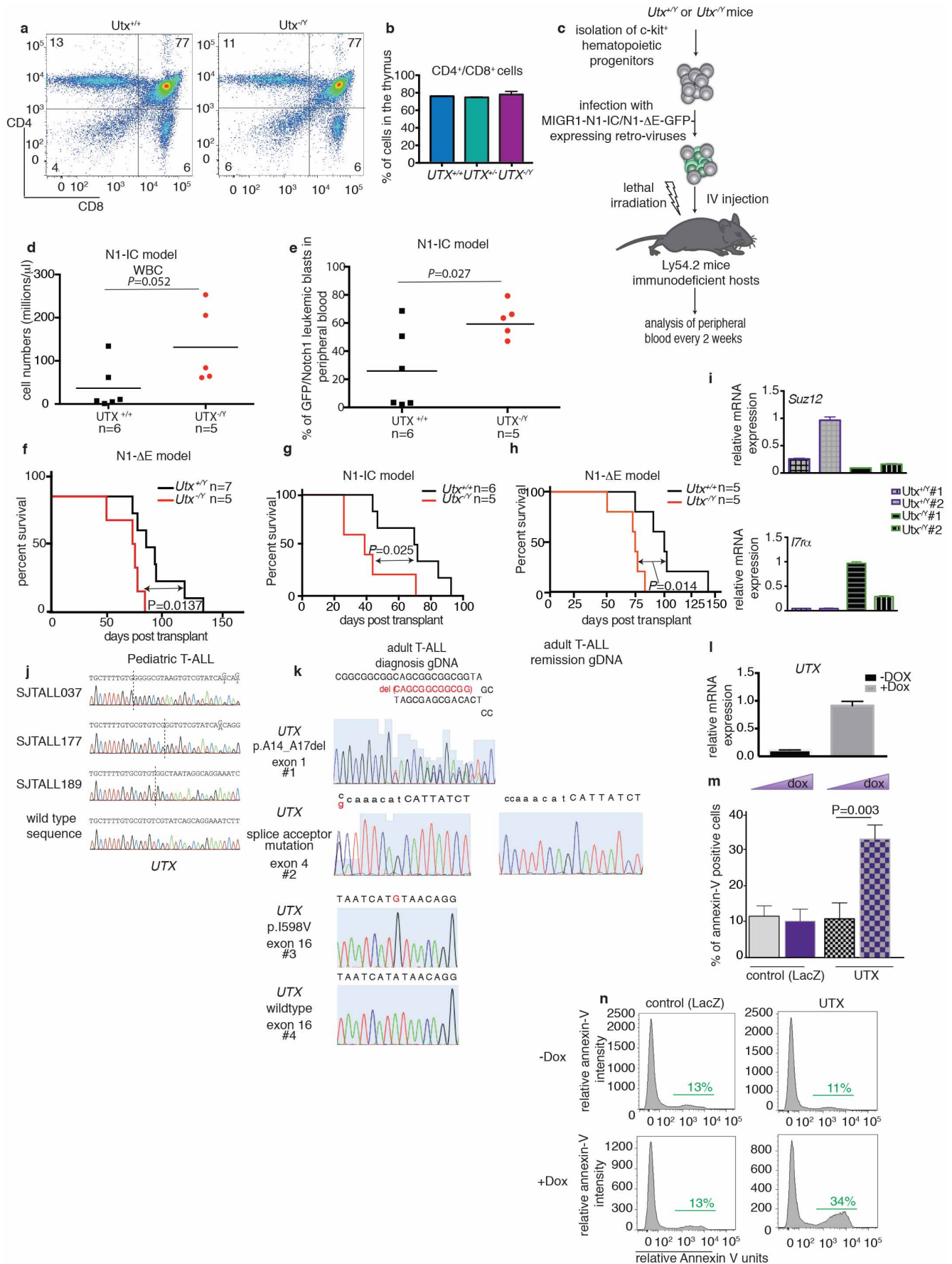
showing the expression levels of important genes in shJMJD3- and shUTX-treated CUTLL1 T-ALL cells. Emphasis is given to the NOTCH1 pathway and apoptosis-related genes. This is a scatter plot representation of an expression analysis comparing three independent studies for shJMJD3 and two for shUTX. **h**, **i**, Scatter plots showing the expression levels of important genes in shJMJD3- and shRenilla-treated CCRF-CEM T-ALL cells (**h**) and in shUTX-treated CCRF-CEM T-ALL cells (**i**). CCRF-CEM cells exhibit increased NOTCH1 levels through mutations in the heterodimerization (HD) domain of NOTCH1 and in the NOTCH1-associated ligase FBXW7. Emphasis is given to the NOTCH1 pathway and apoptosis-related genes. This is a scatter plot representation of an expression analysis comparing two studies for shJMJD3, two for shUTX and two for shRenilla.





**Extended Data Figure 4 | In vivo studies of the role of JMJD3 in T-ALL using luciferase analysis of CEM-, P12- and CUTLL1-based xenograft models in immunocompromised (NRG) mouse recipients.** **a, b,** *In vivo* growth of CEM T-ALL cells in subcutaneous xenograft studies upon genomic ablation of *JMJD3* and *UTX* (red and green circles denote shJMJD3-expressing cells (two different shRNAs); blue denotes shUTX-expressing cells; and black circles denote shRenilla-expressing cells). One million CEM cells were injected into the animals, and representative graphs from five mouse recipients and an image of a representative mouse on days 0 and 6 are shown (**a**). Representative graphs from five mouse recipients and the average luciferase intensity on days 0 and 6 are shown (**b**). **c,** Results for growth of CEM cells at different time points post transplantation in subcutaneous xenograft studies ( $n = 5$ ). **d,** Comparison of *in vivo* cell growth in the subcutaneous model of shJMJD3-, shUTX- and shRenilla-expressing P12 cells ( $n = 5$ ).

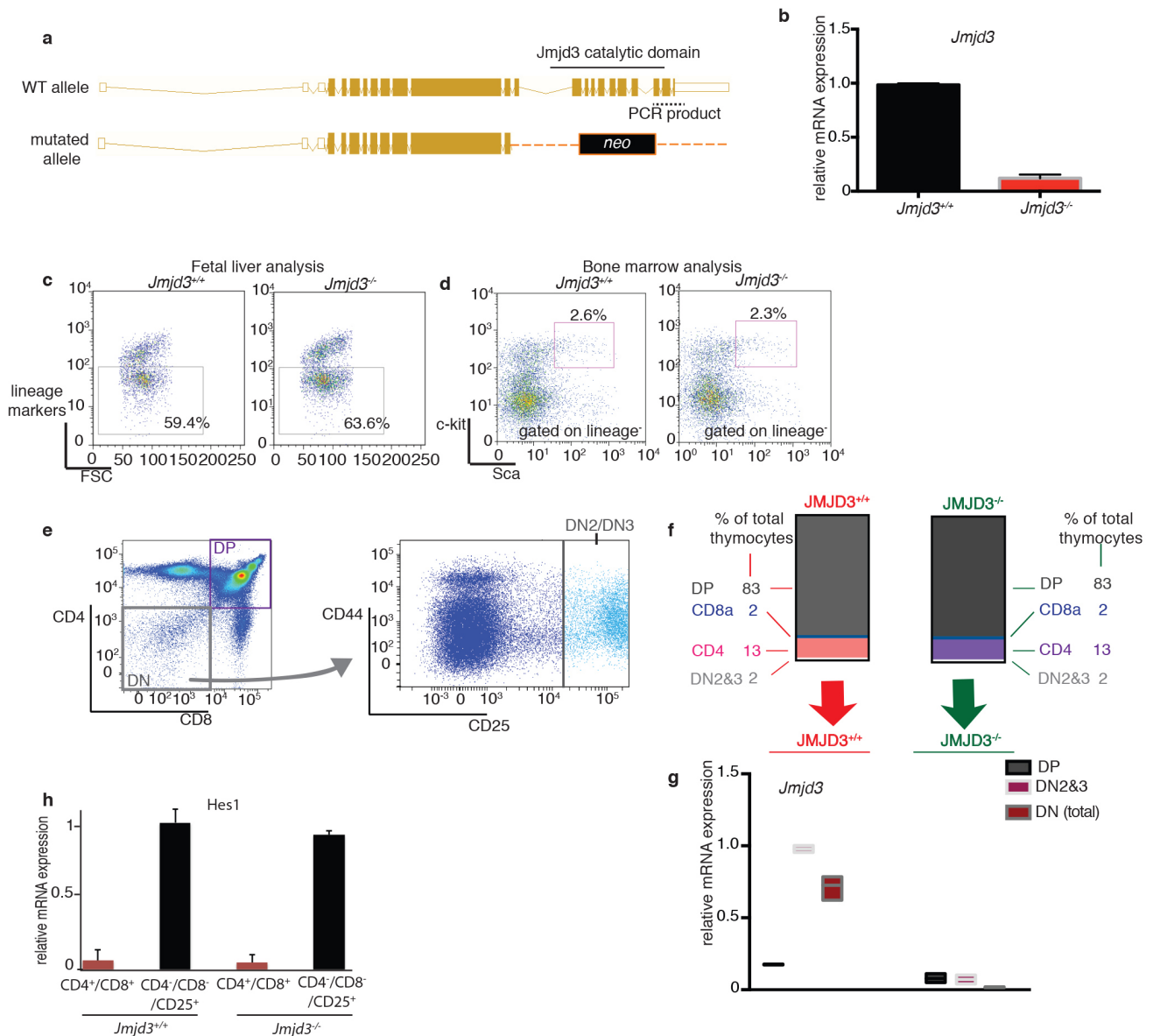
One million P12 cells were injected into sublethally irradiated NRG (immunocompromised) recipients, and the mice were monitored every day for luciferase activity. Day 0 was the first day that a substantially detectable luciferase intensity was measured. The last day of the experiment was the day that either luciferase intensity reached saturation or the mice were euthanized for humanitarian reasons. Red and green circles denote shJMJD3-expressing cells (two different shRNAs, shJMJD3A and shJMJD3B); blue denotes shUTX-expressing cells; and black circles denote shRenilla-expressing cells. **e,** Monitoring the change in luciferase intensity over a period of seven days in the subcutaneous xenograft model using CUTLL1 T-ALL cells ( $n = 4$ ). **f, g,** Intravenous xenograft studies using CUTLL1 cells injected into sublethally irradiated NRG (immunocompromised) recipients ( $n = 8$  or 6, as indicated in the figure). In **e-g**,  $0.5 \times 10^6$  CUTLL1 cells were transplanted, and the mice were monitored every day for luciferase activity.



# Extended Data Figure 5 | UTX is a tumour suppressor and is genetically inactivated in T-ALL but is dispensable for physiological T-cell development.

**a, b**, Study of lymphoid development in  $Utx^{-/Y}$  compared with  $Utx^{+/+}$  (or  $Utx^{+/Y}$ , data not shown) background mice. Flow cytometric analyses of  $CD4^{+}$  and  $CD8^{+}$  expression (**a**), and the relative proportions of  $CD4^{+}CD8^{+}$  (double-positive) thymocytes across different genotypes (**b**) are shown. A representative example from three independent samples (biological replicates) is shown. **c**, Illustration of the transplantation scheme for the *in vivo* leukaemia studies. **d, e**, T-ALL progresses faster in the male  $Utx$  knockout background ( $Utx^{-/Y}$ ) than in the female wild-type background ( $Utx^{+/+}$ ) in recipients transplanted with NOTCH1-IC-GFP-expressing haematopoietic progenitors, as is demonstrated by the white blood cell counts in the peripheral blood (**d**), as well as the percentage of  $GFP^{+}$  leukaemic cells in the peripheral blood upon transplantation of wild-type progenitors (**e**) from female mice ( $Utx^{+/+}$ ) compared with the corresponding knockout cells ( $Utx^{-/Y}$ ). **f**, Survival study of the recipients of cells from male wild-type ( $Utx^{+/Y}$ ,  $n = 7$ ) and knockout ( $Utx^{-/Y}$ ,  $n = 5$ ) mice expressing NOTCH1-deltaE( $\Delta E$ )-GFP (an allele with weaker oncogenic action than NOTCH1-IC). **g, h**, Survival analysis

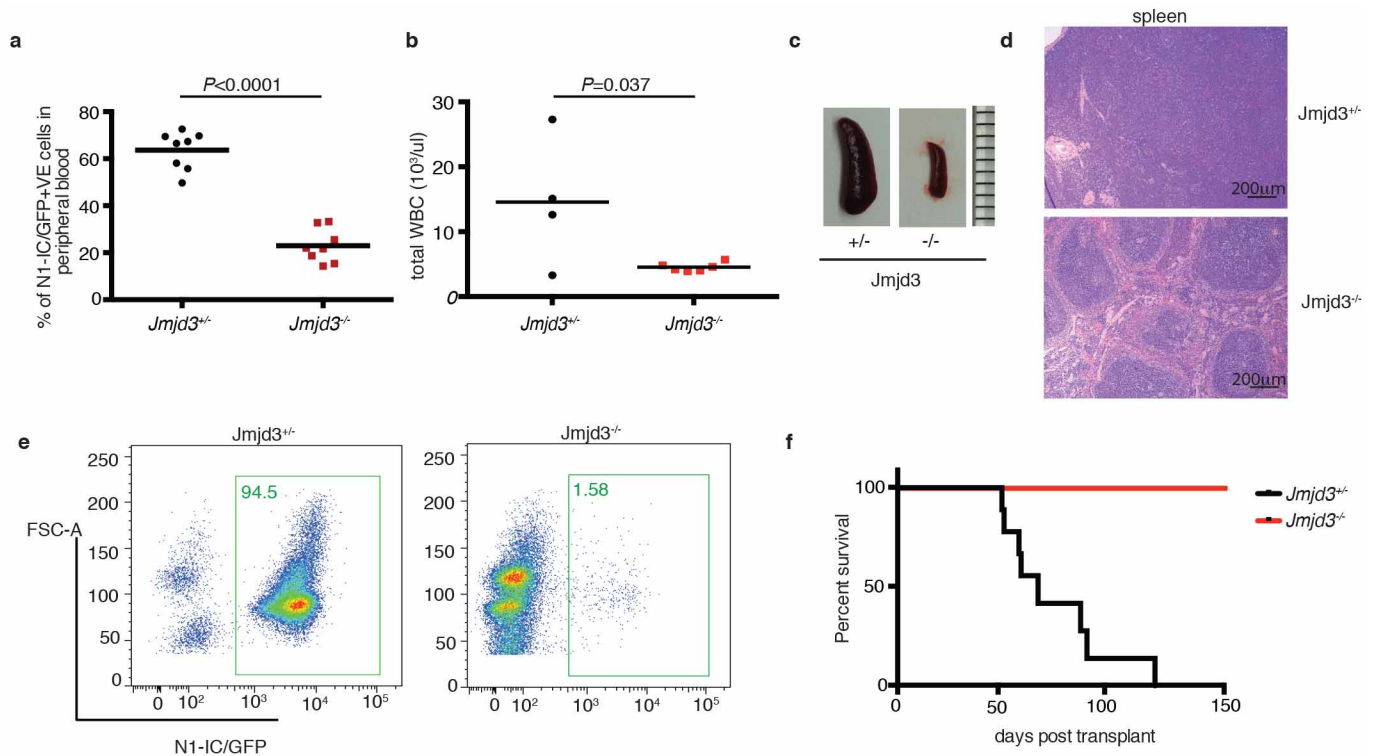
of recipients upon transplantation of wild-type progenitors from female mice ( $Utx^{+/+}$ ) compared with the corresponding knockout cells ( $Utx^{-/Y}$ ) carrying NOTCH1-IC (**g**) or NOTCH1- $\Delta E$  (**h**). **i**, Quantitative PCR (qPCR) validation of the expression levels of one downregulated gene (*Suz12*) and one upregulated gene (*Il7r*) in  $Utx^{-/Y}$  (compared with  $UTX^{+/Y}$ ) mice. The average results from three independent samples (studies) are presented. **j**, Targeted Sanger sequencing in paediatric T-ALL led to the identification of three cases with frameshift mutations. The positions of the mutations are indicated by dashed lines in the electropherograms. **k**, Identification of one in-frame deletion (p.Ala14\_Ala17del, #1, top panel), one splice acceptor site (#2, second panel) and one missense mutation (#3, third panel) in adult T-ALL. Case #4 is an adult T-ALL case with wild-type *UTX* (control, bottom panel). Mutations are indicated by red characters. **l**, The levels of *UTX* in CUTLL1 T-ALL cells in the absence (–dox) or presence (+dox) of doxycycline. **m, n**, Apoptosis analysis through measuring annexin V staining using control LacZ-expressing and *UTX*-expressing CUTLL1 cells in the absence or presence of doxycycline. Representative plots (**l, n**), as well as the average results (**l, m**), from three independent experiments are shown.



**Extended Data Figure 6 | Physiological development of the haematopoietic system in the absence of JMJD3.** **a, b**, Targeting scheme for the generation of the *Jmjd3*<sup>-/-</sup> mouse (**a**) and PCR-based quantification of the wild-type and mutant transcripts (**b**) using a specific primer set for the 3' end of *Jmjd3* cDNA. **c, d**, Analysis of the fetal liver for lineage markers (**c**), as well as the bone marrow (**d**) of recipients for haematopoietic progenitors (the Lin<sup>-</sup> c-Kit<sup>+</sup> Sca1<sup>+</sup> (LSK) population), for the *Jmjd3*<sup>+/+</sup> and *Jmjd3*<sup>-/-</sup> genotypes. Representative plots from three independent experiments are shown.

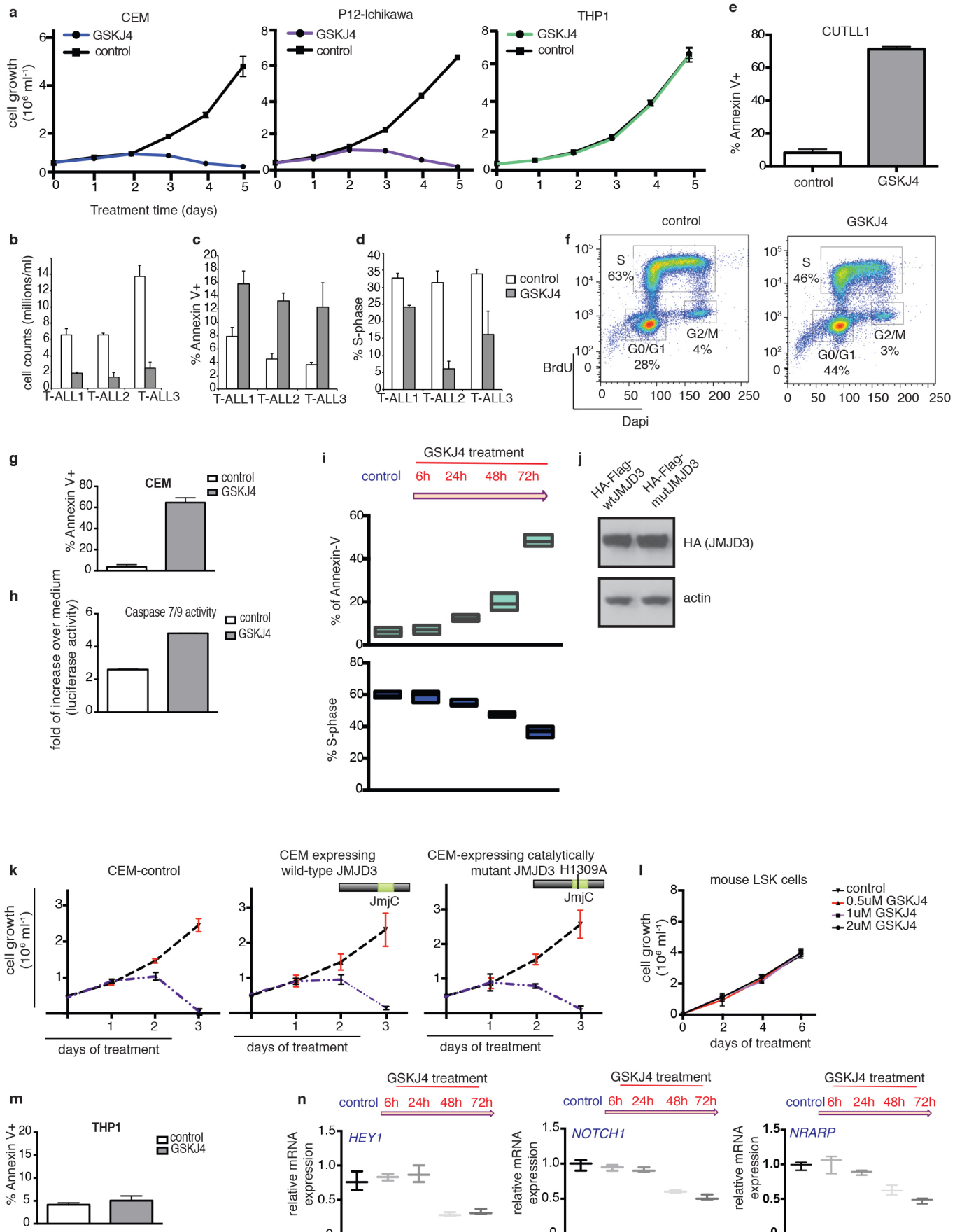
**e–g**, Analysis of major thymic subsets in *Jmjd3*<sup>+/+</sup> ( $n = 7$ ) and *Jmjd3*<sup>-/-</sup> ( $n = 7$ ) mice. Schematic representation of the flow cytometric analysis performed (**e**). Relative proportions of the major cell populations in the thymi of *Jmjd3*<sup>+/+</sup> and *Jmjd3*<sup>-/-</sup> mice (**f**). The mRNA expression of the *Jmjd3* gene at different stages of thymic development (**g**). **h**, The expression of NOTCH1 target genes (such as *Hes1*,  $n = 7$ ) in CD4<sup>+</sup>CD8<sup>+</sup> (double positive) and CD4<sup>-</sup>CD8<sup>-</sup>CD25<sup>+</sup> lymphocyte progenitor cells. Representative plots (**e**), as well as average results (**g, h**), from seven independent thymi are shown.





**Extended Data Figure 7 | JMJD3 is necessary for disease initiation in an animal model of T-ALL.** Initiation of the disease was studied by transplanting c-Kit<sup>+</sup> haematopoietic progenitors. **a, b**, Blood analysis of the recipients for NOTCH1-IC-GFP leukaemic blasts ( $Jmjd3^{+/-}$ ,  $n = 8$ ;  $Jmjd3^{-/-}$ ,  $n = 8$ ; **a**) and white blood cells (WBCs,  $Jmjd3^{+/-}$ ,  $n = 4$ ;  $Jmjd3^{-/-}$ ,  $n = 6$ ; **b**).

**c–e**, Comparison of the organ size (**c**), histochemistry (haematoxylin and eosin staining) (**d**) and flow-cytometry-based analysis (**e**) of the leukaemic cell infiltration of the spleen. **f**, Survival studies of recipients. Eight recipients for the  $Jmjd3^{+/-}$  and eight for the  $Jmjd3^{-/-}$  background were used in **c–f**.



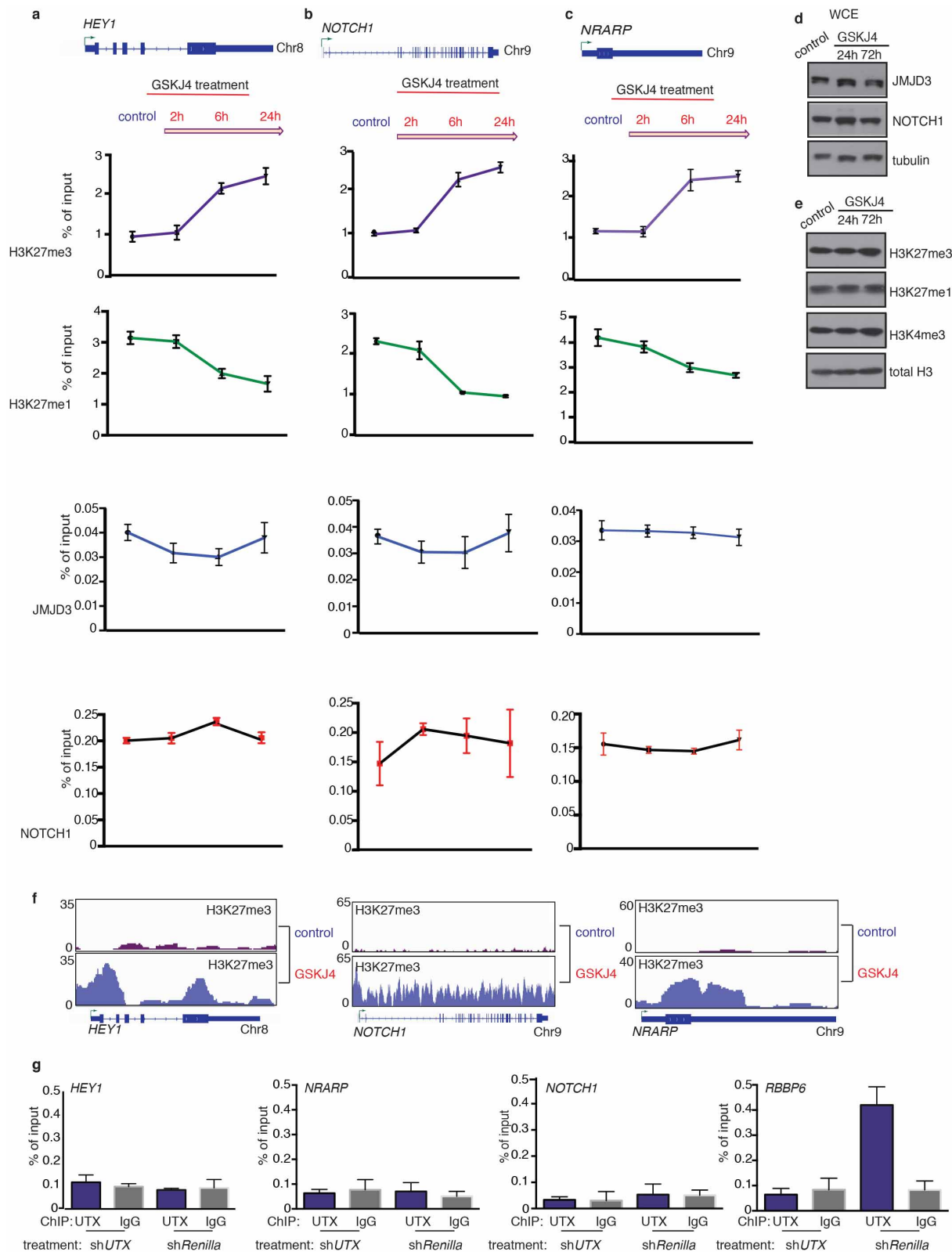
# Extended Data Figure 8 | GSKJ4 inhibitor induces apoptosis and cell cycle arrest of T-ALL but not myeloid leukaemia or physiological LSK cells.

**a**, Effect of GSKJ4 (at 2  $\mu$ M concentration) on a panel of T-ALL and myeloid lines. The average results from three representative studies are shown.

**b–d**, Effects on cell growth (**b**), apoptosis (**c**) and the cell cycle (**d**) in three primary T-ALL lines. The average results from three representative studies are shown. **e, f**, Measurement of apoptosis (**e**,  $n = 3$ ) and cell cycle effects

(**f**, representative study from three experiments) on CUTLL1 cells 72 h post treatment with the inhibitor. **g, h**, Apoptosis assays using annexin V staining of CEM cells (**g**) after a period of 72 h of treatment and measuring caspase 7/9 activity upon treatment of CUTLL1 T-ALL cells with GSKJ5 or GSKJ4 over a period of 24 h (**h**). **i**, Time course studies of annexin V (top) and cell cycle analysis (bottom) of CUTLL1 cells over a period of 72 h during GSKJ4

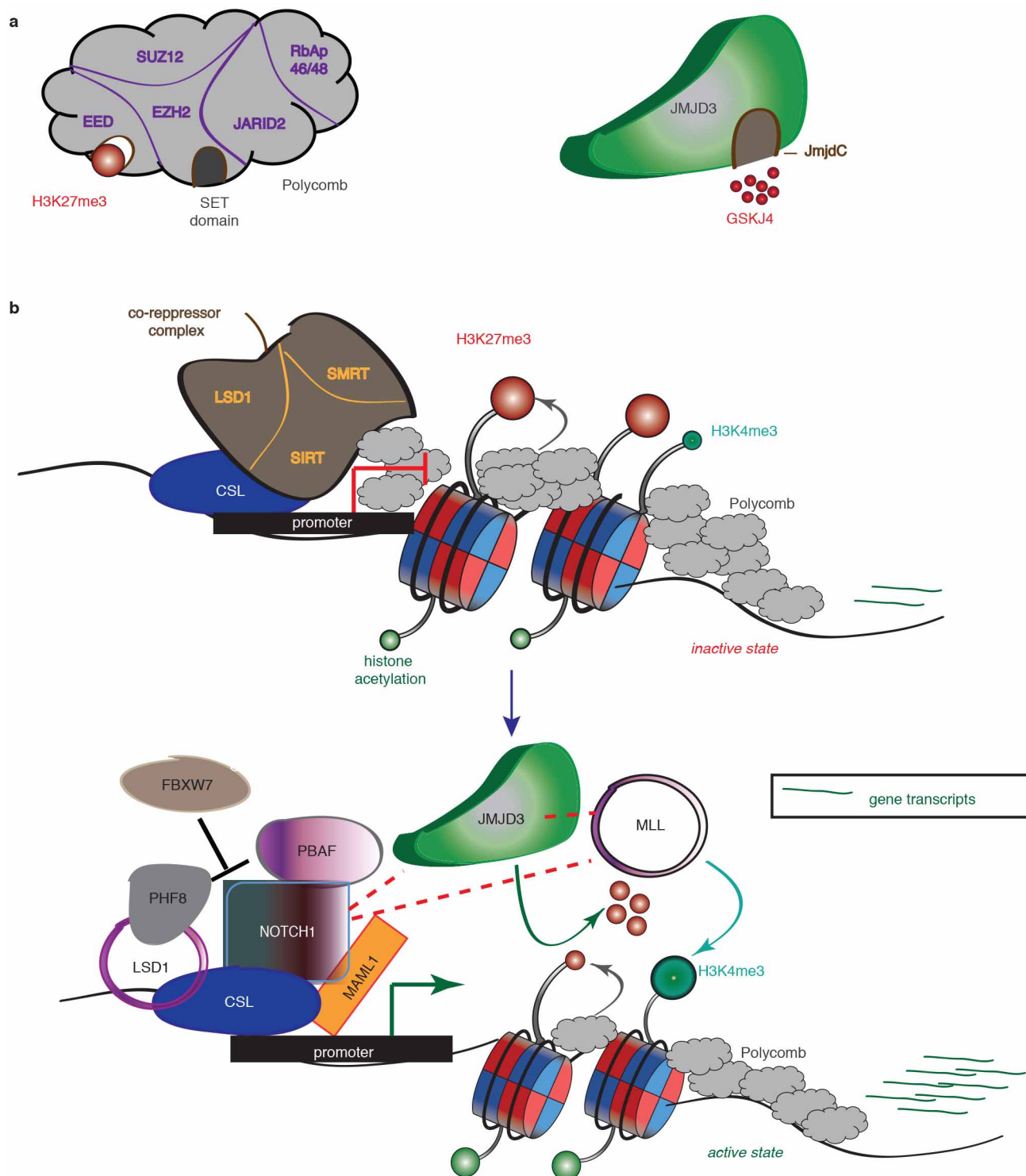
treatment according to the scheme on top of the figure. **j**, Expression of the wild-type and catalytic mutant of JMJD3 in T-ALL (CEM) cells. **k**, Cell growth analysis of T-ALL cells overexpressing wild-type JMJD3 or a catalytic mutant of JMJD3 upon GSKJ4 treatment over a period of 72 h. Average results from three independent experiments are shown. **l**, Cell growth of LSK cells upon treatment with the control (2  $\mu$ M) and different concentrations of the inhibitor GSKJ4. **m**, Annexin V staining of THP-1 (AML) cells after a period of 72 h of GSKJ4 or GSKJ5 (control) treatment at 2  $\mu$ M concentration. The average results from three independent experiments are shown. **n**, The mRNA levels are shown for three classical NOTCH1 targets (*HEY1*, *NRARP* and *NOTCH1*) over a period of 72 h during GSKJ4 treatment. The average results from three independent experiments are shown.



**Extended Data Figure 9 | GSKJ4 treatment leads to increased H3K27me3 levels on NOTCH1 target genes through specific inhibition of JMJD3 activity.** **a–c**, Analysis of the promoter area of *HEY1* (**a**), *NOTCH1* (**b**) and *NRARP* (**c**) for H3K27me3, H3K27me1, NOTCH1 and JMJD3 enrichment over a period of 24 h during GSKJ4 treatment. The average results from three independent experiments are shown. **d**, Analysis of the total protein extracts from CUTLL1 cells for JMJD3 and NOTCH1. **e**, Analysis of the chromatin

fraction from CUTLL1 cells for the repressive mark H3K27me3, the activating marks H3K27me1 and H3K4me3, as well as total histone H3 levels. Representative plots from three independent experiments are shown. **f**, Snapshots of GSKJ4-associated H3K27me3 changes in major NOTCH1 and JMJD3 targets. **g**, ChIP-qPCR analyses for UTX binding to the NOTCH1 target genes *HEY1*, *NRARP* and *NOTCH1*. (*RBBP6* was used as positive control). The average results from three independent experiments are shown.





**Extended Data Figure 10 | JMJD3 as a pivotal factor in NOTCH1-mediated oncogenic activation in T-cell leukaemia.** **a**, Schematic representation of the H3K27me3 writer (the polycomb complex, left) and eraser (JMJD3, right). EZH2 contains the catalytic subunit of the complex through its SET domain, whereas the EED subunit recognizes the H3K27me3 mark and aids in polycomb binding. JmjdC domain activity is inhibited by the small molecule inhibitor GSKJ4. **b**, The main idea about the key role of JMJD3 in the NOTCH1

transcriptional complex. Before activation of the NOTCH1 signalling pathway, the promoters of classical NOTCH1 target genes are bound by RBP-J $\kappa$ , together with components of the co-repressor complexes and PRC2, leading to low gene expression. After the binding of NOTCH1 and its co-activator MAML1, the genes are activated through the recruitment of JMJD3 and the MLL complex, with simultaneous eviction of PRC2, which leads to the demethylation of H3K27me3 and the methylation of H3K4me3.

# Structure and mechanism of $\text{Zn}^{2+}$ -transporting P-type ATPases

Kaituo Wang<sup>1†\*</sup>, Oleg Sitsel<sup>1\*</sup>, Gabriele Meloni<sup>1</sup>, Henriette Elisabeth Autzen<sup>1</sup>, Magnus Andersson<sup>2</sup>, Tetyana Klymchuk<sup>1</sup>, Anna Marie Nielsen<sup>1</sup>, Douglas C. Rees<sup>3</sup>, Poul Nissen<sup>1</sup> & Pontus Gourdon<sup>1†</sup>

**Zinc is an essential micronutrient for all living organisms. It is required for signalling and proper functioning of a range of proteins involved in, for example, DNA binding and enzymatic catalysis<sup>1</sup>. In prokaryotes and photosynthetic eukaryotes,  $\text{Zn}^{2+}$ -transporting P-type ATPases of class IB (ZntA) are crucial for cellular redistribution and detoxification of  $\text{Zn}^{2+}$  and related elements<sup>2,3</sup>. Here we present crystal structures representing the phosphoenzyme ground state (E2P) and a dephosphorylation intermediate (E2·P<sub>i</sub>) of ZntA from *Shigella sonnei*, determined at 3.2 Å and 2.7 Å resolution, respectively. The structures reveal a similar fold to  $\text{Cu}^{+}$ -ATPases, with an amphipathic helix at the membrane interface. A conserved electronegative funnel connects this region to the intramembranous high-affinity ion-binding site and may promote specific uptake of cellular  $\text{Zn}^{2+}$  ions by the transporter. The E2P structure displays a wide extracellular release pathway reaching the invariant residues at the high-affinity site, including C392, C394 and D714. The pathway closes in the E2·P<sub>i</sub> state, in which D714 interacts with the conserved residue K693, which possibly stimulates  $\text{Zn}^{2+}$  release as a built-in counter ion, as has been proposed for  $\text{H}^{+}$ -ATPases. Indeed, transport studies in liposomes provide experimental support for ZntA activity without counter transport. These findings suggest a mechanistic link between P<sub>IB</sub>-type  $\text{Zn}^{2+}$ -ATPases and P<sub>III</sub>-type  $\text{H}^{+}$ -ATPases and at the same time show structural features of the extracellular release pathway that resemble P<sub>II</sub>-type ATPases such as the sarcoplasmic/endoplasmic reticulum  $\text{Ca}^{2+}$ -ATPase<sup>4,5</sup> (SERCA) and  $\text{Na}^{+}$ ,  $\text{K}^{+}$ -ATPase<sup>6</sup>. These findings considerably increase our understanding of zinc transport in cells and represent new possibilities for biotechnology and biomedicine.**

Zinc is an abundant transition metal in life, serving multiple functions<sup>1</sup>, yet elevated concentrations of  $\text{Zn}^{2+}$  are toxic, as are its heavy-metal mimetics such as  $\text{Cd}^{2+}$  and  $\text{Pb}^{2+}$  (ref. 7).  $\text{Zn}^{2+}$ -transporting P-type ATPases (the P<sub>IB</sub>-ATPases ZntA and CadA) are active transporters that are crucial for the cellular detoxification of these elements<sup>3</sup>, as well as for the subcellular redistribution of micronutritional zinc<sup>2</sup>. The significance of  $\text{Zn}^{2+}$ -ATPases is further underscored by the presence of multiple and occasionally redundant genes encoding these enzymes in higher plants such as *Arabidopsis thaliana*<sup>2</sup>. The lack of ZntA in animals, the prevalence of such enzymes in pathogens, and the fact that zinc is exploited in the host–microorganism arms race (for example, to inactivate vital virulence determinants of *Streptococcus pneumoniae*<sup>8</sup>) make these P<sub>IB</sub>-ATPases attractive targets for new antibiotics, antifungals and herbicides. ZntA couples ATP hydrolysis at the intracellular A (actuator/dephosphorylation), P (phosphorylation) and N (nucleotide binding) domains to ion efflux through the M (transmembrane) domain (Extended Data Fig. 1a). The mechanism is schematically described by the ‘Post–Albers’ cycle<sup>9</sup>, which has four principal states (E1, E1P, E2P and E2) that define alternating access to an intramembranous high-affinity ion-binding site<sup>10</sup> (Fig. 1a, centre, and Extended Data Fig. 1b). However, the only structures

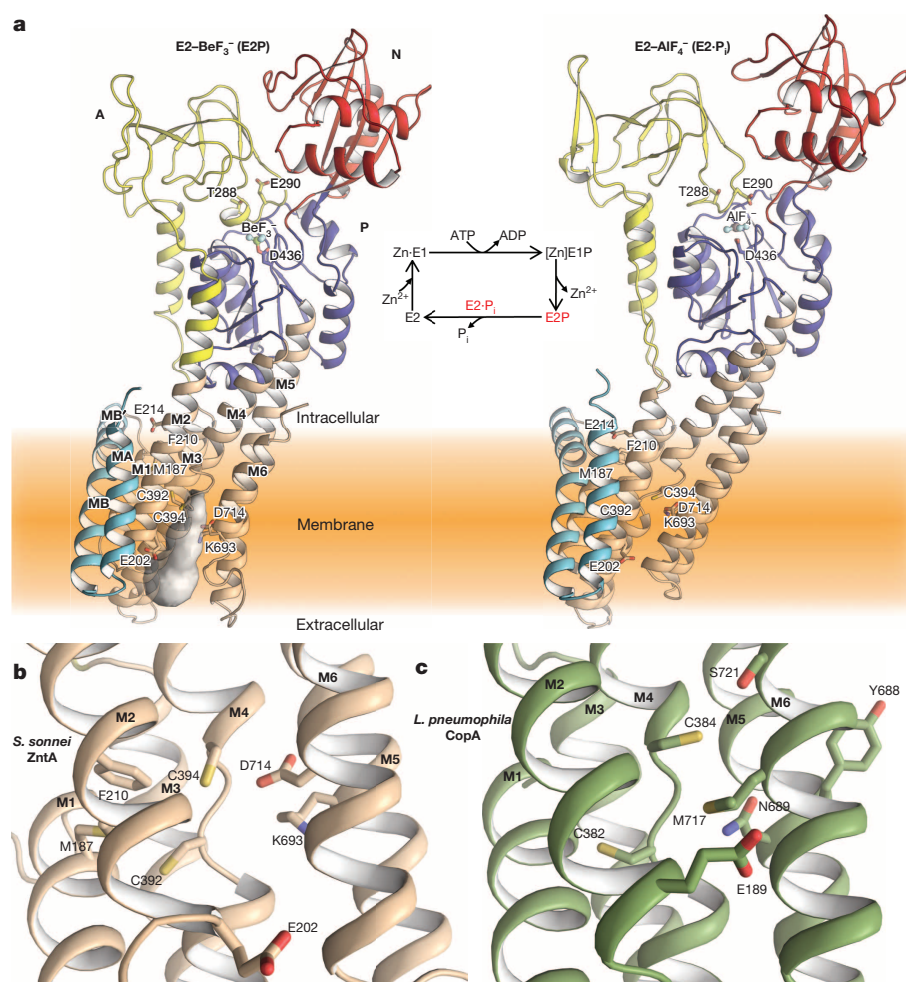
that have been determined for this class of protein are for the related  $\text{Cu}^{+}$ -transporting P<sub>IB</sub>-ATPase CopA<sup>11,12</sup> and for a ZntA domain<sup>13</sup>, limiting the functional and mechanistic understanding of this class of proteins. Fundamental questions that remain to be answered include how zinc transport is accomplished across the membrane and coupled to ATPase activity and how sequence motifs that are specific to  $\text{Zn}^{2+}$ -ATPases relate to structure and function.

We have determined the crystal structures of two reaction cycle intermediates of ZntA from *S. sonnei*, which is 99.2% identical to the *Escherichia coli* ZntA (the best characterized member of the family) and is stimulated by the equivalent ions *in vitro* (Fig. 1a and Extended Data Figs 2–4a). Crystals were obtained using a modified HiLiDe (high concentrations of lipid and detergent) technique<sup>14</sup> (see Methods) and in the presence of the zinc chelator TPEN (*N,N,N',N'*-tetrakis(2-pyridinylmethyl)-1,2-ethanediamine) plus either  $\text{BeF}_3^-$  or  $\text{AlF}_4^-$ , mimicking the zinc-free phosphoenzyme ground state (denoted E2P) and a dephosphorylation intermediate (E2·P<sub>i</sub>), respectively. The structures were determined at 3.2 Å and 2.7 Å resolution (Extended Data Table 1) and reveal a P<sub>IB</sub>-type ATPase fold reminiscent of CopA, with intracellular A, P and N domains and eight similarly arranged transmembrane segments (MA, MB and M1–M6), albeit with shorter extracellular loops (Extended Data Fig. 5). The heavy-metal binding domain (HMBD), a characteristic feature of P<sub>IB</sub>-ATPases (Extended Data Fig. 1a), was, however, not visible in the electron density maps, as was also the case for CopA<sup>11</sup>. The intracellular domains are arranged differently in the two *S. sonnei* ZntA structures, in agreement with the equivalent states of CopA and SERCA<sup>4,11,12</sup>:  $\text{BeF}_3^-$  mimics the phosphorylation of D436 (*S. sonnei* ZntA numbering throughout), which is buried and protected by the catalytic TGE loop of the A domain in this E2P-like state, whereas the TGE motif activates a water molecule coordinated to  $\text{AlF}_4^-$ , imitating dephosphorylation at D436 as in an E2·P<sub>i</sub>-like state (Extended Data Fig. 6).

The single intramembranous high-affinity  $\text{Zn}^{2+}$ -binding site of ZntA<sup>10</sup> deserves particular attention. Biochemical studies have indicated that  $\text{Zn}^{2+}$  binding depends on C392 and C394 (in the CPC motif of the M4 segment), K693 (M5) and D714 (M6)<sup>10,15</sup>. In the structures, these four residues overlap well with the equivalent cysteines, asparagine and methionine in the corresponding E2·P<sub>i</sub> state of  $\text{Cu}^{+}$ -ATPases<sup>11</sup> (Fig. 1b, c). Further supporting an important functional role of these four residues, the only other conserved side chains in the region that may participate in  $\text{Zn}^{2+}$  binding are those of M187 and Y354, but our mutations of these residues do not affect function (Fig. 2a). However, the K693 side chain would be an unexpected ligand for  $\text{Zn}^{2+}$  (refs 16, 17), and indeed  $\text{Zn}^{2+}$  binding is unaffected by the mutation of lysine to alanine at this position (K693A) (Fig. 2b), suggesting that binding is instead established by the two cysteine thiolates and two oxygen ligands, possibly from D714 in a bidentate fashion, which is a recurrent coordination motif of  $\text{Zn}^{2+}$ -binding sites<sup>16,17</sup>. Congruent with this role, the relative activity of the

<sup>1</sup>Centre for Membrane Pumps in Cells and Disease (PUMPKin), Danish National Research Foundation, Aarhus University, Department of Molecular Biology and Genetics, Gustav Wieds Vej 10C, DK-8000 Aarhus C, Denmark. <sup>2</sup>Science for Life Laboratory, Department of Theoretical Physics, Swedish e-Science Research Center, KTH Royal Institute of Technology, SE-171 21 Solna, Sweden. <sup>3</sup>Division of Chemistry and Chemical Engineering and Howard Hughes Medical Institute, California Institute of Technology, 1200 East California Boulevard, Pasadena, California 91125, USA. <sup>†</sup>Present addresses: Department of Biomedical Sciences, University of Copenhagen, Blegdamsvej 3B, DK-2200 Copenhagen, Denmark (K.W. and P.G.); Department of Experimental Medical Science, Lund University, Solvegatan 19, SE-221 84 Lund, Sweden (P.G.).

\*These authors contributed equally to this work.



**Figure 1 | Structures of the *S. sonnei*  $\text{Zn}^{2+}$ -ATPase.** **a**, The  $\text{E2-BeF}_3^-$  ( $\text{E2P}$ , left) and  $\text{E2-AlF}_4^-$  ( $\text{E2}\cdot\text{P}_i$ , right) structures with class-specific helices MA, MB and MB' coloured in cyan, helices M1–M6 coloured in beige, and the A, P and N domains coloured in yellow, blue and red, respectively (domain names are highlighted in bold). Key residues for function are highlighted. An extracellular release pathway (white surface) is present only in the  $\text{E2P}$  state, as computed with CAVER<sup>30</sup>. A schematic Post–Albers reaction cycle<sup>9</sup> for ZntA is shown (centre) with the experimentally determined structures marked in red. **b**, Close view of the intramembranous ion-binding region coloured as in **a**, displaying the proposed  $\text{Zn}^{2+}$ -binding residues C392 and C394 (in M4), K693 (M5) and D714 (M6)<sup>10,15</sup>. **c**, Equivalent view to **b** of the *Legionella pneumophila*  $\text{Cu}^+$ -ATPase CopA (Protein Data Bank (PDB) ID, 3RFU<sup>11</sup>). Critical CopA residues overlie equally important residues of  $\text{Zn}^{2+}$ -ATPases (see also Extended Data Fig. 2). Side chain atoms are depicted in blue (nitrogen) and red (oxygen).

D714E mutant decreases with the increasing ionic radii and coordination distances of  $\text{Zn}^{2+}$ ,  $\text{Cd}^{2+}$  and  $\text{Pb}^{2+}$  (Fig. 2a).

What therefore is the role of the essential K693? One striking difference between the ion-binding region (between M4, M5 and M6) of CopA and ZntA is the aforementioned D714 in *S. sonnei* ZntA. The side chain of D714 is stabilized by K693 in the  $\text{E2}\cdot\text{P}_i$  state (Fig. 3a and Extended Data Fig. 7a): this interaction potentially has important functional implications, with the charge-stabilizing lysine residue possibly acting as a built-in counter ion in zinc-free states (that is, as observed here). Such a mechanism was proposed earlier for plasma membrane  $\text{H}^+$ -ATPases<sup>18</sup> (Fig. 3c). Indeed, residues R655 and D684, which form this pair in the *Arabidopsis thaliana*  $\text{H}^+$ -ATPase AHA2, are located at positions that almost overlap the positions of K693 and D714 of *S. sonnei* ZntA<sup>18</sup>, pointing to common principles of ion transport in  $\text{P}_{\text{IB}}^-$  and  $\text{P}_{\text{III}}^-$  type ATPases.

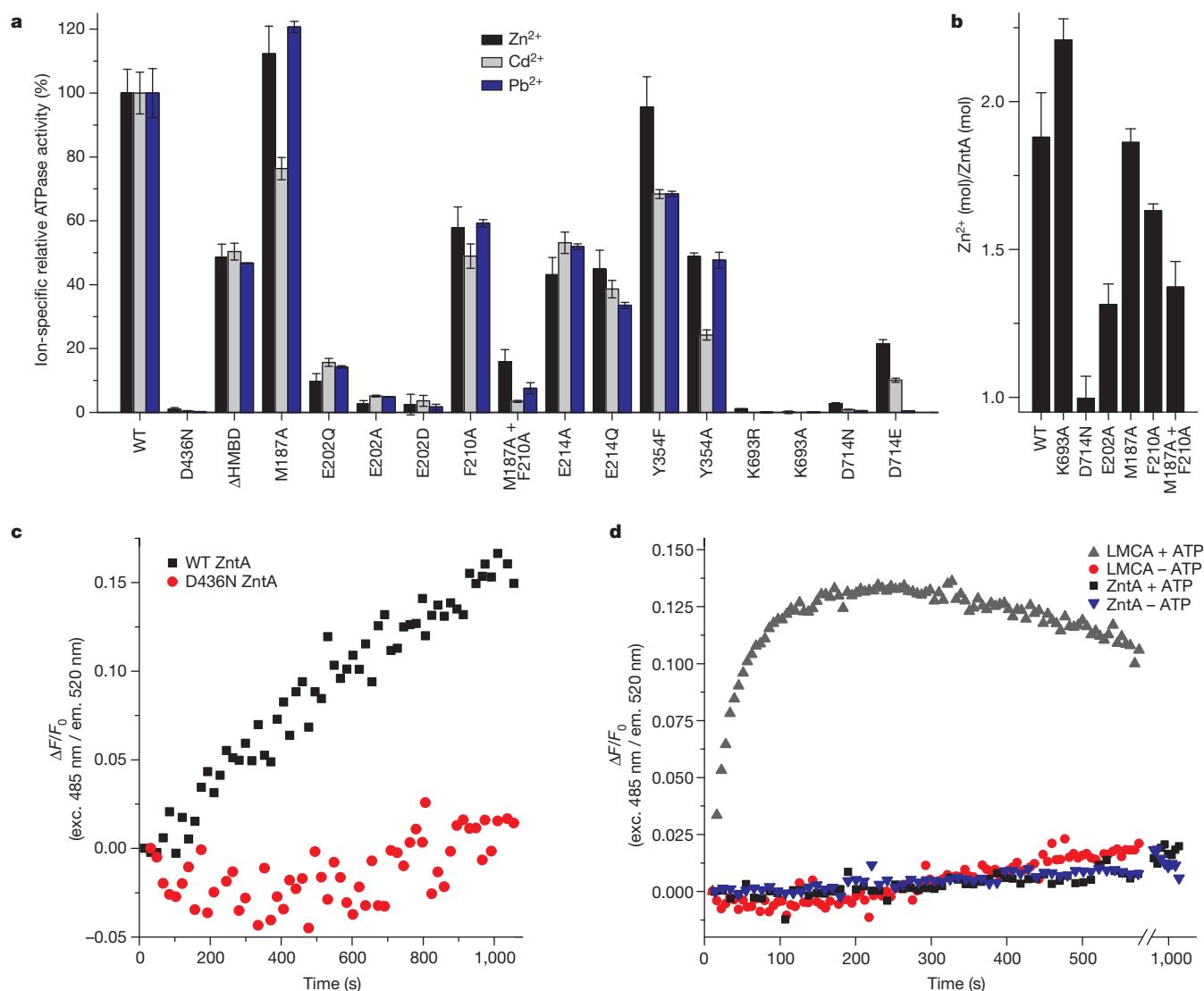
Transport and putative  $\text{H}^+$  counter transport were then analysed in proteoliposomes. Whereas  $\text{Zn}^{2+}$  accumulated in vesicles (Fig. 2c), we were unable to detect any changes in intravesicular pH (Fig. 2d). As a positive control, we used the  $\text{Ca}^{2+}$ -ATPase LMCA1 from *Listeria monocytogenes*, which showed clear  $\text{H}^+$  antiporter activity<sup>19</sup> (Fig. 2d). Furthermore, while the electron density maps allowed the identification of several water molecules in the  $\text{E2-AlF}_4^-$  structure, no sites were detected that could be ascribed to, for example,  $\text{K}^+$ ,  $\text{Na}^+$ ,  $\text{Ca}^{2+}$  or  $\text{Mg}^{2+}$  counter ions, and these cations were also not required for ZntA activity (Extended Data Fig. 4b and see Methods for details). All considered, our observations thus support zinc flux without associated counter-ion transport.

In  $\text{Cu}^+$ -ATPases, ion release has been proposed to occur via a pathway lined by MA, M2 and M6 that remains open in the  $\text{E2P}$  and  $\text{E2}\cdot\text{P}_i$  states<sup>12</sup>. We were consequently surprised to find that no extracellular

pathway was evident in the  $\text{E2}\cdot\text{P}_i$  state of *S. sonnei* ZntA, in contrast to CopA, and that, instead, substantial conformational changes occurred in the M domain in the  $\text{E2P}$  to  $\text{E2}\cdot\text{P}_i$  state transition. These conformational changes resemble, by contrast, those of SERCA, in which a wide opening appears between M1–M2, M3–M4 and M5–M6 in the  $\text{E2P}$  state<sup>4,5</sup> and reseals in the occluded  $\text{E2}\cdot\text{P}_i$  state (Fig. 3d and Extended Data Fig. 1). In ZntA, the extracellular portions of M5–M6 shift away from the  $\text{Zn}^{2+}$ -binding CPC motif, and rearrangements (less pronounced than those of SERCA) in M2 and M3–M4 expose the high-affinity site to the extracellular side (Fig. 3d and Extended Data Fig. 7a). This SERCA-like pathway must allow release of free zinc into the extracellular environment, as further supported by an observed reorientation of the sulphur side chains of the CPC motif away from the ion-binding site between the  $\text{E2P}$  and  $\text{E2}\cdot\text{P}_i$  states. With K693 being flexible without a strong interaction with D714 in the  $\text{E2P}$  state (Fig. 3b), it is possible that K693 has an additional role: electrostatic repulsion against the re-entry of  $\text{Zn}^{2+}$ , possibly further stimulated by E202 guiding  $\text{Zn}^{2+}$  into the extracellular environment. The equivalent residue to E202 in SERCA and CopA (E90 and E189, respectively) has been proposed to serve a similar purpose<sup>11,20</sup>, and supporting this notion, E202 is critical for enzyme function<sup>21</sup> (Fig. 2a). Furthermore, E202 showed considerable conformational flexibility in a 60-ns molecular dynamics simulation of the open  $\text{E2P}$  structure, linking the intramembranous ion-binding site to the extracellular environment, as is also supported by steered molecular dynamics simulations of  $\text{Zn}^{2+}$  passage from the CPC motif to the extracellular environment (Extended Data Fig. 7b–e).

One important consideration is how  $\text{Zn}^{2+}$  is initially delivered to ZntA from the intracellular milieu. Although the current structures of  $\text{Zn}^{2+}$ -free states are outward-oriented and therefore closed towards





**Figure 2 | Functional studies of zinc, cadmium, lead and counter-ion transport in *S. sonnei* ZntA.** **a**, ATP turnover associated with different *S. sonnei* ZntA constructs in detergent-lipid solution, relative to wild-type activity with each ion. The specific activities of wild-type (WT) *S. sonnei* ZntA with  $\text{Zn}^{2+}$ ,  $\text{Cd}^{2+}$  and  $\text{Pb}^{2+}$  were  $592 \pm 23$ ,  $491 \pm 10$  and  $813 \pm 23$   $\text{nmol P}_i \text{mg}^{-1} \text{min}^{-1}$ , respectively; the mean + s.d. of technical replicates is shown ( $n = 3$ ). **b**,  $\text{Zn}^{2+}$  binding to different *S. sonnei* ZntA constructs, as determined using the dye Zincon. ZntA binds to two  $\text{Zn}^{2+}$  ions: one binds to

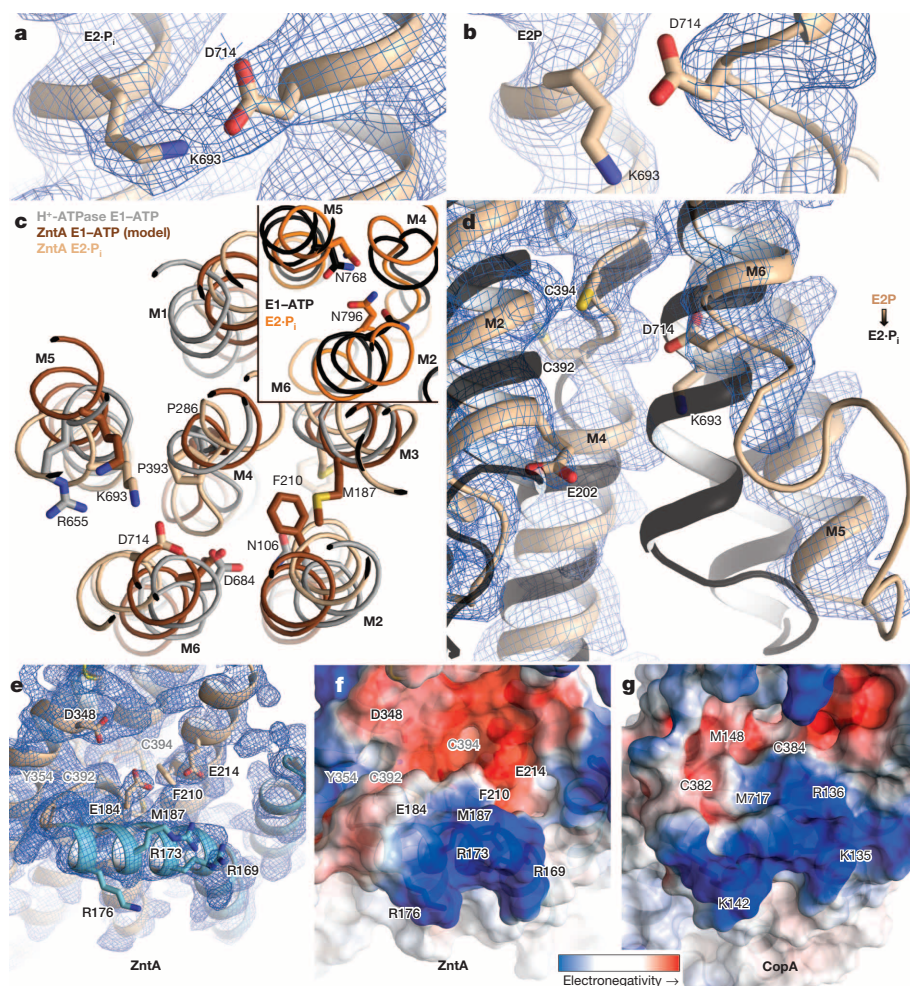
the high-affinity site in the transmembrane domain, and one binds to the HMBD. The mean + s.d. of biological replicates is shown ( $n = 3$ ). **c**, **d**, Zinc transport by wild-type and D436N *S. sonnei* ZntA proteoliposomes (**c**) and measurements of  $\text{H}^+$  counter-ion transport by wild-type *S. sonnei* ZntA and  $\text{Ca}^{2+}$ -ATPase LMCA proteoliposomes (**d**), as monitored using the zinc-selective chelator FluoZin-1 (**c**) and the pH indicator pyranine (**d**) (see also Extended Data Fig. 4c and Methods). exc., excitation; em., emission.

the intracellular side (Extended Data Fig. 1b), they hint at how  $\text{Zn}^{2+}$  entry may take place. The uptake of intracellular cations by P-type ATPases is expected to occur at the membrane interface at M1 (refs 4, 5, 11, 18, 22, 23), and in CopA through an entry site with an invariant methionine. Sequence analyses show that M1 segments in  $\text{Zn}^{2+}$ -ATPases also harbour a conserved methionine (M187), although this residue is located closer to the CPC motif in ZntA (Fig. 1b, c and Extended Data Fig. 2), but mutational studies indicate that this residue alone is not essential (Fig. 2a). However, in contrast to CopA, the entry area in  $\text{Zn}^{2+}$ -ATPases displays a conserved and negatively charged funnel structure (lined by E184, E214 and D348 at the membrane interface) that stretches towards the intramembranous ion-binding site and that is plugged by M187 and F210, the latter of which is conserved as a phenylalanine or tyrosine in ZntA (Fig. 3e, f, Extended Data Fig. 8a and Extended Data Table 2). Whereas the activity of the F210A mutant is only moderately affected *in vitro*, the M187A and F210A double mutant is inactive, with less zinc binding than the wild type (Fig. 2a, b). We note that the equivalent

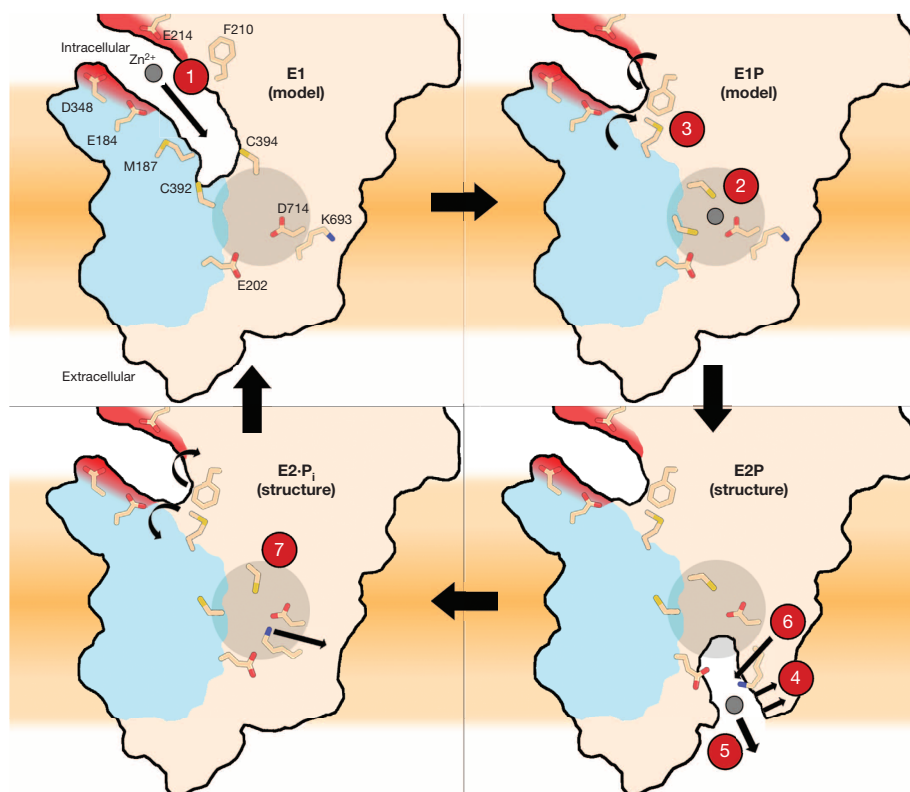
residue to *S. sonnei* ZntA F210 in  $\text{H}^+$ -ATPases, N106, is a gatekeeper for  $\text{H}^+$  entry<sup>18,24</sup> (Fig. 3c). With the conformational changes anticipated for the shift to the E1 states,  $\text{Zn}^{2+}$  may thus be guided by M187 through the funnel and led directly to C392 in the high-affinity site, which is capped by M148 and F210. Because the funnel is narrow and negatively charged, we find it likely that free  $\text{Zn}^{2+}$  ions and not a glutathione-ligated complex will interact with the funnel, unlike the proposed uptake mechanism of the heavy-metal ABC exporter Atm1 from *Novosphingobium aromaticivorans*<sup>25</sup>.

The role of the HMBD of P<sub>IB</sub>-ATPases is puzzling<sup>11,26,27</sup>. In  $\text{Cu}^+$ -ATPases, a platform formed by an amphipathic helix, MB', at the intracellular membrane interface (Fig. 3g) has been proposed to serve as an interaction site for HMBDs, as well as for metal-donating chaperones, allowing allosteric regulation and copper supply to the ATPase core. The MB' platform and its amphipathic character are maintained in ZntA, exposing several positively charged residues to the intracellular side (Fig. 3e, f and Extended Data Table 2). However, as no equivalent chaperones





**Figure 3 | Details of the *S. sonnei* ZntA structures.** The blue mesh represents the final  $2F_o - F_c$  electron density, contoured at  $1\sigma$  (other colours as in Fig. 1a unless noted). See Methods for additional details on figures. **a**, Close view of K693 and D714 in the E2·P<sub>i</sub> state. **b**, Close view of K693 and D714 in the E2P state. **c**, Comparison of the transmembrane regions of *S. sonnei* ZntA–AlF<sub>4</sub><sup>–</sup> in the E2·P<sub>i</sub> state, *A. thaliana* AHA2 H<sup>+</sup>–ATPase (in the E1–ATP state; PDB ID, 3B8C<sup>18</sup>; grey) and an E1–ATP model of *S. sonnei* ZntA (brown). Inset, identical view of the equivalent region of the E1–ATP (black) and E2·P<sub>i</sub> (orange) states of SERCA (PDB ID, 1T5S and 3B9R, respectively). **d**, Structural differences between the extracellular portions of the E2P (coloured as in Fig. 1a) and the E2·P<sub>i</sub> structures (black) (see also Extended Data Fig. 7a). **e–g**, The MB' platform of the E2·P<sub>i</sub> state of *S. sonnei* ZntA (**e**, **f**) and *L. pneumophila* CopA (**g**).



**Figure 4 | Putative zinc transport mechanism of ZntA.** A transport cycle based on schematic models of the E1 and E1P states and the E2P and E2·P<sub>i</sub> structures. In the presence of intracellular zinc, Zn<sup>2+</sup> enters the ATPase through the electronegative funnel (red) at the MB' platform (1). Upon Zn<sup>2+</sup> binding (2) to the intramembranous ion-binding site (grey circle), F210 and M187 occlude the ion entry funnel (3), preventing backflow of Zn<sup>2+</sup>. Substantial domain rearrangements in transition to the E2P state open the extracellular pathway (4), lowering the affinity for Zn<sup>2+</sup> and mediating Zn<sup>2+</sup> release (5), possibly stimulated by K693 (6). Dephosphorylation triggers closure of the transmembrane domain, in which K693 (as a built-in counter ion) forms a salt bridge with D714. Upon dephosphorylation, the side chains move to their initial positions (7) before an E2 to E1 transition is stimulated by the presence of intracellular Zn<sup>2+</sup>.

are known for zinc, the metal is most likely delivered by chelators such as glutathione, rendering the HMBD the most likely interaction candidate for the MB' platform in ZntA. Using a known structure of the almost identical HMBD of *E. coli* ZntA<sup>13</sup>, the ClusPro docking server docks the domain immediately at MB', stabilized by charge complementation (as was also proposed for CopA<sup>27,28</sup>) with the metal-binding CXXC motif (where X denotes any amino acid residue) being solvent accessible in the vicinity of the entry funnel (Extended Data Fig. 8b–d). Truncations and mutations of the HMBD retain a functional ZntA, only with reduced activity<sup>15,29</sup> (Fig. 2a), and we therefore favour an autoregulatory role for this domain.

The first atomic structures of a Zn<sup>2+</sup>-transporting P<sub>IB</sub>-type ATPase reveal unique features. These include an intracellular, negatively charged and presumably ion-catching funnel, a high-affinity Zn<sup>2+</sup>-binding site with a putative lysine switch acting as a built-in counter ion (with an unexpected similarity to P<sub>III</sub>-type plasma membrane H<sup>+</sup>-ATPases) and an extracellular Zn<sup>2+</sup>-release pathway (which, unlike that of copper-transporting P<sub>IB</sub>-type ATPases, resembles that of the classical P<sub>II</sub>-type ion pumps). These findings significantly increase our understanding of zinc transport in cells (Fig. 4) and represent new possibilities for biotechnology and biomedicine. Detailed insight into the transport mechanism and specificity determinants may, for example, aid in using plant biotechnology to accumulate valuable zinc in edible plants or to decontaminate heavy metals in soil, and the release pathway may be a favourable target site for new antibiotics.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 22 January; accepted 25 June 2014.

Published online 17 August 2014.

- Berg, J. M. & Shi, Y. The galvanization of biology: a growing appreciation for the roles of zinc. *Science* **271**, 1081–1085 (1996).
- Williams, L. E. & Mills, R. F. P. P<sub>IB</sub>-ATPases—an ancient family of transition metal pumps with diverse functions in plants. *Trends Plant Sci.* **10**, 491–502 (2005).
- Argüello, J. M., Gonzalez-Guerrero, M. & Raimunda, D. Bacterial transition metal P<sub>IB</sub>-ATPases: transport mechanism and roles in virulence. *Biochemistry* **50**, 9940–9949 (2011).
- Olesen, C. *et al.* The structural basis of calcium transport by the calcium pump. *Nature* **450**, 1036–1042 (2007).
- Toyoshima, C., Norimatsu, Y., Iwasawa, S., Tsuda, T. & Ogawa, H. How processing of aspartylphosphate is coupled to luminal gating of the ion pathway in the calcium pump. *Proc. Natl Acad. Sci. USA* **104**, 19831–19836 (2007).
- Laursen, M., Yatime, L., Nissen, P. & Fedosova, N. U. Crystal structure of the high-affinity Na<sup>+</sup>K<sup>+</sup>-ATPase-ouabain complex with Mg<sup>2+</sup> bound in the cation binding site. *Proc. Natl Acad. Sci. USA* **110**, 10958–10963 (2013).
- Domingo, J. L. Metal-induced developmental toxicity in mammals: a review. *J. Toxicol. Environ. Health* **42**, 123–141 (1994).
- Couñago, R. M. *et al.* Imperfect coordination chemistry facilitates metal ion release in the Psa permease. *Nature Chem. Biol.* **10**, 35–41 (2014).
- Albers, R. W. Biochemical aspects of active transport. *Annu. Rev. Biochem.* **36**, 727–756 (1967).
- Raimunda, D., Subramanian, P., Stemmler, T. & Argüello, J. M. A tetrahedral coordination of zinc during transmembrane transport by P-type Zn<sup>2+</sup>-ATPases. *Biochim. Biophys. Acta* **1818**, 1374–1377 (2012).
- Gourdon, P. *et al.* Crystal structure of a copper-transporting P<sub>IB</sub>-type ATPase. *Nature* **475**, 59–64 (2011).
- Andersson, M. *et al.* Copper-transporting P-type ATPases use a unique ion-release pathway. *Nature Struct. Mol. Biol.* **21**, 43–48 (2014).
- Banci, L. *et al.* A new zinc-protein coordination site in intracellular metal trafficking: solution structure of the Apo and Zn(II) forms of ZntA(46–118). *J. Mol. Biol.* **323**, 883–897 (2002).
- Gourdon, P. *et al.* HiLiDe-systematic approach to membrane protein crystallization in lipid and detergent. *Cryst. Growth Des.* **11**, 2098–2106 (2011).
- Okkeri, J. & Haltia, T. The metal-binding sites of the zinc-transporting P-type ATPase of *Escherichia coli*. Lys693 and Asp714 in the seventh and eighth transmembrane segments of ZntA contribute to the coupling of metal binding and ATPase activity. *Biochim. Biophys. Acta* **1757**, 1485–1495 (2006).
- Patel, K., Kumar, A. & Durani, S. Analysis of the structural consensus of the zinc coordination centers of metalloprotein structures. *Biochim. Biophys. Acta* **1774**, 1247–1253 (2007).
- Andreini, C., Bertini, I. & Cavallaro, G. Minimal functional sites allow a classification of zinc sites in proteins. *PLoS ONE* **6**, e26325 (2011).
- Pedersen, B. P., Buch-Pedersen, M. J., Morth, J. P., Palmgren, M. G. & Nissen, P. Crystal structure of the plasma membrane proton pump. *Nature* **450**, 1111–1114 (2007).
- Faxen, K. *et al.* Characterization of a *Listeria monocytogenes* Ca<sup>2+</sup> pump: a SERCA-type ATPase with only one Ca<sup>2+</sup>-binding site. *J. Biol. Chem.* **286**, 1609–1617 (2011).
- Clausen, J. D. & Andersen, J. P. Glutamate 90 at the luminal ion gate of sarcoplasmic reticulum Ca<sup>2+</sup>-ATPase is critical for Ca<sup>2+</sup> binding on both sides of the membrane. *J. Biol. Chem.* **285**, 20780–20792 (2010).
- Zhitnitsky, D. & Lewinson, O. Identification of functionally important conserved trans-membrane residues of bacterial P<sub>IB</sub>-type ATPases. *Mol. Microbiol.* **91**, 777–789 (2014).
- Winther, A. M. *et al.* The sarcolipin-bound calcium pump stabilizes calcium sites exposed to the cytoplasm. *Nature* **495**, 265–269 (2013).
- Toyoshima, C. *et al.* Crystal structures of the calcium pump and sarcolipin in the Mg<sup>2+</sup>-bound E1 state. *Nature* **495**, 260–264 (2013).
- Ekberg, K., Wielandt, A. G., Buch-Pedersen, M. J. & Palmgren, M. G. A conserved asparagine in a P-type proton pump is required for efficient gating of protons. *J. Biol. Chem.* **288**, 9610–9618 (2013).
- Lee, J. Y., Yang, J. G., Zhitnitsky, D., Lewinson, O. & Rees, D. C. Structural basis for heavy metal detoxification by an Atm1-type ABC exporter. *Science* **343**, 1133–1136 (2014).
- Gonzalez-Guerrero, M. & Argüello, J. M. Mechanism of Cu<sup>+</sup>-transporting ATPases: soluble Cu<sup>+</sup> chaperones directly transfer Cu<sup>+</sup> to transmembrane transport sites. *Proc. Natl Acad. Sci. USA* **105**, 5992–5997 (2008).
- Mattle, D. *et al.* On allosteric modulation of P-type Cu-ATPases. *J. Mol. Biol.* **425**, 2299–2308 (2013).
- Padilla-Benavides, T., McCann, C. J. & Argüello, J. M. The mechanism of Cu<sup>+</sup> transport ATPases: interaction with Cu<sup>+</sup> chaperones and the role of transient metal-binding sites. *J. Biol. Chem.* **288**, 69–78 (2013).
- Dutta, S. J., Liu, J., Stemmler, A. J. & Mitra, B. Conservative and nonconservative mutations of the transmembrane CPC motif in ZntA: effect on metal selectivity and activity. *Biochemistry* **46**, 3692–3703 (2007).
- Chovancova, E. *et al.* CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. *PLoS Comput. Biol.* **8**, e1002708 (2012).

**Acknowledgements** We thank J. L. Karlsen for support with crystallographic computing. O.S. and H.E.A. are supported by the Graduate School of Science and Technology at Aarhus University. G.M. is supported by a Marie Curie International Outgoing Fellowship (European Commission, grant no. 252961). M.A. was supported by a Marie Curie Career Integration Grant (FP7-MC-CIG-618558). P.N. was supported by an advanced research grant from the European Research Council (250322 Biomemos), and P.G. was supported, in the later stage, by the Lundbeck Foundation and the Swedish Research Council (K2013-99X-22251-01-5). We are grateful for assistance with crystal screening from Maxlab, beam lines 911-2/3, and with data collection from the Swiss Light Source, beam line X06SA. Access to synchrotron sources was supported by the Dancscatt program of the Danish Council of Independent Research and by BioStruct-X contract 860.

**Author Contributions** K.W., O.S., T.K. and A.M.N. cloned the *S. sonnei* ZntA constructs. K.W. and O.S. performed protein purification, crystallization and activity measurements in solution. G.M. conducted the vesicle and zinc binding studies, which were developed with D.C.R. K.W. processed the data and solved the crystal structures, and all authors analysed the results. H.E.A. and M.A. conducted molecular dynamics simulations in the absence and presence of zinc, respectively. P.N. and P.G. designed the project. K.W., O.S. and G.M. generated the figures. P.N. and P.G. wrote the paper, and all authors commented on the paper.

**Author Information** Atomic coordinates and structure factors for the *S. sonnei* ZntA (UniProt ID, Q3YW59) E2–AlF<sub>4</sub><sup>–</sup> and E2–BeF<sub>3</sub><sup>–</sup> crystal structures have been deposited in the Protein Data Bank (PDB) under accession numbers 4UMW and 4UMV. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.G. ([pontus@sund.ku.dk](mailto:pontus@sund.ku.dk)).



## METHODS

**Protein expression and purification.** Several ZntA homologues from different prokaryotes were cloned and tested for expression, purification and crystallization in a parallel approach. *S. sonnei* ZntA (UniProt ID, Q3YW59) was cloned into pET-52 with a carboxy-terminal hexahistidine tag and transformed into the C41(DE3) *E. coli* expression strain. Cells were grown in LB medium at 37 °C to an absorbance at 600 nm ( $A_{600}$ ) of 1.0, and the shaker flasks were cooled for 30 min with iced water. Expression was then induced with 1 mM isopropyl- $\beta$ -D-thiogalactoside (IPTG) (final concentration) at 20 °C for 20 h. Harvested cells were resuspended in TKG buffer (17 g cells per 100 ml buffer) containing 20 mM Tris-HCl, pH 7.5, 200 mM KCl and 20% (v/v) glycerol and then frozen at -20 °C. Before cell rupture, the solution was added (final concentrations, 5 mM MgCl<sub>2</sub>, 5 mM  $\beta$ -mercaptoethanol (BME), 2  $\mu$ g ml<sup>-1</sup> DNase I, 1 mM phenylmethanesulphonyl fluoride and Roche protease inhibitor cocktail (1 tablet per 200 ml)), and the cells were lysed using a high pressure homogenizer (three times, 15,000–20,000 p.s.i.). The sample was then kept at 4 °C throughout the entire purification procedure until crystallization. Cell debris was removed by centrifugation at 23,000g for 20 min, and membranes were isolated by ultracentrifugation at 250,000g for 3 h. The membrane pellet was resuspended in 20 mM Tris-HCl, pH 7.5, 200 mM KCl, 20% (v/v) glycerol, 5 mM MgCl<sub>2</sub> and 5 mM BME, to a final concentration of 12 ml buffer per g membrane and then exposed to 10 mg ml<sup>-1</sup> (final concentration) octaethylene glycol monododecyl ether (C<sub>12</sub>E<sub>8</sub>) for 1 h with gentle stirring. Unsolubilized material was removed by ultracentrifugation at 250,000g for 30 min. The supernatant was supplemented with imidazole and solid KCl (final concentrations of 50 and 500 mM, respectively), filtered (0.22  $\mu$ m) and then applied to several sequential 5-ml pre-packed Ni<sup>2+</sup>-chelating columns (HisTrap HP, GE Healthcare; material from 41 cells per column). The columns were washed with buffer containing 20 mM Tris-HCl, pH 7.5, 200 mM KCl, 20% (v/v) glycerol, 5 mM MgCl<sub>2</sub>, 5 mM BME, 0.15 mg ml<sup>-1</sup> C<sub>12</sub>E<sub>8</sub> and 50 mM imidazole until the absorption at 280 nm ( $A_{280}$ ) reached the baseline, and elution was achieved with an additional 450 mM imidazole (final concentration). The *S. sonnei* ZntA-containing fractions were pooled, and the protein was concentrated to 20 mg ml<sup>-1</sup> and then subjected to size-exclusion chromatography. Protein (50 mg) was injected into an XK16/100 column prepared with a 100 ml column volume of Superose 6 Prep Grade (GE Healthcare) equilibrated with 20 mM MOPS-KOH, pH 6.8, 80 mM KCl, 20% (v/v) glycerol, 5 mM MgCl<sub>2</sub>, 5 mM BME and 0.15 mg ml<sup>-1</sup> C<sub>12</sub>E<sub>8</sub>, and the resultant main peak from each run was pooled and concentrated to 12 mg ml<sup>-1</sup>, aliquoted, flash frozen in liquid nitrogen and stored at -80 °C. Yields exceeded 10 mg purified protein per 1 *E. coli* cell culture. The final protein purity was monitored using SDS-PAGE, and the protein concentration was assessed by measuring  $A_{280}$ .

**Crystallization.** *S. sonnei* ZntA aliquots were thawed and supplemented with 4 mg ml<sup>-1</sup> (final concentration) C<sub>12</sub>E<sub>8</sub> and incubated without stirring for 16 h at 4 °C, reaching a modified HiLiDe condition<sup>14</sup>. The sample was then ultracentrifuged at 100,000g for 10 min, diluted to a final concentration of about 6–8 mg ml<sup>-1</sup> protein and treated with 10 mM NaF, 2 mM AlCl<sub>3</sub> or BeSO<sub>4</sub>, 2 mM EGTA and 10  $\mu$ M N,N,N',N'-tetrakis(2-pyridylmethyl)-1,2-ethanediamine (TPEN) (final concentrations) for 30 min. Crystals were grown using the hanging drop vapour diffusion method at 19 °C. The best *S. sonnei* ZntA E2- $\text{AlF}_4^-$  crystals were grown using a reservoir with 300 mM lithium acetate, 3% (v/v) *t*-butanol, 14% polyethylene glycol 2000 monomethyl ether (PEG 2000 MME), 7% (v/v) sorbitol, 10% (v/v) glycerol and 5 mM BME. By contrast, the best *S. sonnei* ZntA E2- $\text{BeF}_3^-$  crystals were obtained using a reservoir with 100 mM MgCl<sub>2</sub>, 200 mM lithium acetate, 17% (v/v) PEG 2000 MME, 10% (v/v) glycerol and 5 mM BME. More than 1,000 crystals were fished with litho-loops, flash cooled in liquid nitrogen and tested at synchrotron sources. The final data sets were collected at the Swiss Light Source, Villigen, Switzerland, using the X06SA beam line and a wavelength of 1.0000 Å (0.9787 Å for Se-E2·P<sub>i</sub>[AlF]), a temperature of 100 K and the Pilatus 6M pixel detector.

**Data processing and structure determination.** Data were processed and scaled with the program XDS<sup>31</sup> to 2.7 Å and 3.2 Å resolution. The E2- $\text{AlF}_4^-$  and E2- $\text{BeF}_3^-$  crystals belonged to space groups C22<sub>2</sub> and P2<sub>1</sub>, respectively. Initial phases for the E2- $\text{AlF}_4^-$  form were obtained with molecular replacement using Phaser<sup>32</sup>, and monomer A of *L. pneumophila* CopA (PDB ID, 3RFU<sup>11</sup>) was used as a search model. Anomalous peaks in a Se-SAD (Se single-wavelength anomalous diffraction) data set of the E2- $\text{AlF}_4^-$  form were calculated using the molecular replacement phases, and the Se-Met positions were used to guide model building. Model building was performed with Coot<sup>33</sup>, using the *L. pneumophila* CopA structure as a template. Model refinement was carried out using phenix.refine<sup>34</sup>, applying TLS parameters in the late stages of refinement only, reaching  $R_{\text{cryst}}/R_{\text{free}}$  values of 20.7/24.0 (E2- $\text{AlF}_4^-$ ) and 21.0/28.1 (E2- $\text{BeF}_3^-$ ). The E2- $\text{BeF}_3^-$  form was also determined by molecular replacement using the refined *S. sonnei* ZntA E2- $\text{AlF}_4^-$  structure as a search model and refined using a similar procedure. The final refinement statistics are listed in Extended Data Table 1. Structures were analysed using MolProbity, indicating that 96.35/94.42, 3.49/5.08 and 0.17/0.51% of the residues were in the

favoured, allowed and non-favoured regions, with 6.64/10.91% rotamer outliers and 6.48/10.11% as clash scores, respectively, for the E2·P<sub>i</sub> and E2P states.

**Functional characterization.** The purification protocol for functionally assessed *S. sonnei* ZntA constructs was similar to the one described for crystallization ( $\Delta$ HMBD lacks the first 103 residues). However, following affinity chromatography, the samples were treated with 1 mM EDTA and then subjected to a 5-ml HiTrap desalting column (GE Healthcare) using the equivalent SEC buffer to that for crystallization. Release of inorganic phosphate (P<sub>i</sub>) associated with ATPase activity was assessed using the Baginski assay<sup>35</sup>. The reaction system contained 5  $\mu$ g protein, 40 mM MOPS-KOH, pH 6.8, 150 mM NaCl, 5 mM KCl, 5 mM MgCl<sub>2</sub>, 3.0 mg ml<sup>-1</sup> C<sub>12</sub>E<sub>8</sub>, 1.2 mg ml<sup>-1</sup> soybean lipid, 20 mM cysteine, 5 mM NaN<sub>3</sub> and 0.25 mM Na<sub>2</sub>MoO<sub>4</sub> in a total volume of 50  $\mu$ l. This solution was first incubated with different transition metal ions or EGTA, supplemented with 3 mM ATP (final concentration) to start the reaction and then incubated for 10 min while shaking at 37 °C. Freshly prepared stop solution (50  $\mu$ l) (2.5% (w/v) ascorbic acid, 0.4 M (v/v) HCl, 0.48% (w/v) (NH<sub>4</sub>)<sub>2</sub>MoO<sub>4</sub> and 0.8% SDS) was then added to stop the reaction and start colour development. After 10 min incubation at 18 °C, 75  $\mu$ l colorimetric solution (2% (w/v) arsenite, 2% (v/v) acetic acid and 3.5% (w/v) sodium citrate) was added to the mixture and incubated for another 30 min at 18 °C. Absorbance was measured at 860 nm. One experiment with three replicates was performed for each construct and ion.

**Reconstitution in proteoliposomes.** *E. coli* polar lipids (25 mg ml<sup>-1</sup>) and egg-yolk phosphatidylcholine (25 mg ml<sup>-1</sup>) in chloroform were mixed at a 3:1 (w/w) ratio and dried under a nitrogen stream and continuous rotation to form a homogeneous thin film in a glass balloon. Lipids were desiccated overnight under vacuum (protected from light) and suspended in 1 mM dithiothreitol (DTT) to a final concentration of 25 mg ml<sup>-1</sup>. A concentrated stock (10 $\times$ ) was used to bring the suspension to a final concentration of 20 mM MOPS, pH 6.8, 250 mM NaCl and 1 mM DTT. Lipids were subjected to three rounds of freeze-thawing in liquid nitrogen. Proteoliposomes were prepared by extrusion (11 times) through 0.2- $\mu$ m polycarbonate filters to form large unilamellar vesicles (LUVs) using a mini extruder (Avanti Polar Lipids) equipped with two 1-ml gas-tight syringes. Proteoliposomes were destabilized by the addition of *n*-dodecyl- $\beta$ -D-maltoside (DDM) to a final concentration of 0.02% (w/v) and tilting for 1 h at 18 °C and were subsequently placed on ice for 10 min. Wild-type and D436N *S. sonnei* ZntA, as well as LMCA1, were added (1–2 mg ml<sup>-1</sup>, purified essentially as described for crystallization) to a final protein-to-lipid ratio of 1:20 (w/w), and the mixture was incubated for 1 h at 4 °C under tilting. Control liposomes were prepared using the same procedure without the addition of protein. Detergent was removed through consecutive incubations with activated Bio-Beads SM-2 (Bio-Rad), by exchanging the beads after 1, 16, 18 and 20 h. The Bio-Beads were subsequently removed, and the proteoliposomes were collected by ultracentrifugation at 163,000g for 45–60 min at 4 °C and resuspended in 20 mM MOPS, pH 6.8, 250 mM NaCl and 1 mM DTT (Buffer PL) to a final protein concentration of 1 mg ml<sup>-1</sup>.

**Zinc transport assays using FluoZin-1.** Wild-type and D436N *S. sonnei* ZntA proteoliposomes were diluted 1:2 in 20 mM MOPS, pH 6.8, 250 mM NaCl and 1 mM DTT to a protein concentration of 0.5 mg ml<sup>-1</sup>. A stock of the fluorescent Zn<sup>2+</sup> chelator FluoZin-1 (2 mM in H<sub>2</sub>O) was added to a final concentration of 200  $\mu$ M. FluoZin-1 encapsulation was performed by three freeze-thaw cycles and subsequent extrusion through 0.2- $\mu$ m polycarbonate filters. Proteoliposomes were collected by ultracentrifugation at 163,000g for 45–60 min at 4 °C, and the supernatant containing excess FluoZin-1 was removed. Proteoliposomes were washed with 1 ml Buffer PL, collected by ultracentrifugation and suspended in the same buffer (1 ml). Transport assays were performed in the presence of a final concentration of 10 mM MgCl<sub>2</sub> on 100  $\mu$ l samples. The reactions were initiated by the addition of concentrated stocks of ZnCl<sub>2</sub> (1 mM) and ATP (10 mM) stock to final concentrations of 40  $\mu$ M ZnCl<sub>2</sub> and 1 mM ATP. A fluorescence time course was measured in a 96-well plate reader using an excitation wavelength of 485 nm and an emission wavelength of 520 nm. Experiments in the absence of ATP were performed in parallel as controls. The ATP-dependent Zn<sup>2+</sup> transport was determined as  $\Delta F/F_0$ , where  $\Delta F$  is the difference between the fluorescence measured in the presence and the absence of ATP, and  $F_0$  is the fluorescence recorded immediately after ATP addition. Each condition was tested at least in duplicate, and one representative trace is shown for each.

**H<sup>+</sup> counter-ion transport assays using pyranine.** *S. sonnei* ZntA (wild type) and LMCA1 and control proteoliposomes were diluted 1:2 to final concentration of 20 mM MOPS, pH 7.0, 250 mM NaCl, 100 mM KCl, 10 mM MgCl<sub>2</sub> and 1 mM DTT (Buffer Counter). A stock of the fluorescent pH indicator pyranine (0.1 M in H<sub>2</sub>O) was added to a final concentration of 10 mM. Pyranine encapsulation was performed using three freeze-thaw cycles and subsequent extrusion through 0.2- $\mu$ m polycarbonate filters. Proteoliposomes were collected by ultracentrifugation at 163,000g for 45–60 min at 4 °C, and the supernatant was removed. Proteoliposomes were washed with 1 ml Buffer Counter, collected by ultracentrifugation and suspended in the same buffer. The reactions were initiated by the addition of concentrated stocks

of  $\text{ZnCl}_2$  (1 mM) or  $\text{CaCl}_2$  (2.5 mM) and ATP (10 mM) to obtain a final concentration of 40  $\mu\text{M}$   $\text{ZnCl}_2$  (for *S. sonnei* ZntA) or 100  $\mu\text{M}$   $\text{CaCl}_2$  (for LMCA1), and 1 mM ATP. Experiments in the absence of ATP were performed in parallel, as well as experiments on control liposomes. A fluorescence time course was measured in a 96-well plate reader using an excitation wavelength of 450 nm and an emission wavelength of 520 nm. The ATP-dependent  $\text{H}^+$  counter-ion transport was determined as  $\Delta F/F_0$ , where  $\Delta F$  is the difference between the fluorescence measured in proteoliposomes and in control liposomes and  $F_0$  is the fluorescence recorded immediately after ATP addition. Each condition was tested at least in duplicate, and one representative trace is shown for each.

**Effect of  $\text{Na}^+$  or  $\text{K}^+$  on *S. sonnei* ZntA activity.** To investigate the effect of  $\text{Na}^+$  or  $\text{K}^+$  on the activity of wild-type *S. sonnei* ZntA in detergent micelles, the buffer in *S. sonnei* ZntA stock solutions was exchanged using 5-ml HiTrap desalting columns packed with Sephadex G-25 resin with a  $\text{K}^+$ -depleted solution (20 mM MOPS, pH 6.8, 250 mM NaCl, 1 mM DTT, 0.01  $\text{mg ml}^{-1}$   $\text{C}_{12}\text{E}_8$  and 20% (v/v) glycerol) or a  $\text{Na}^+$ -depleted solution (20 mM MOPS, pH 6.8, 250 mM KCl, 1 mM DTT, 0.01  $\text{mg ml}^{-1}$   $\text{C}_{12}\text{E}_8$  and 20% (v/v) glycerol). To exchange the buffer in proteoliposome preparations, 100  $\mu\text{l}$  stocks were diluted in 1 ml 20 mM MOPS, pH 6.8, 250 mM NaCl and 1 mM DTT or 20 mM MOPS, pH 6.8, 250 mM KCl and 1 mM DTT and regenerated by using three freeze–thaw cycles. Proteoliposomes were collected by ultracentrifugation, washed with 1 ml of the corresponding buffer and regenerated by using three freeze–thaw cycles. This procedure was repeated, and the proteoliposomes were subsequently extruded through 0.2- $\mu\text{m}$  polycarbonate filters. Proteoliposomes were collected by ultracentrifugation and suspended in a final volume of 100  $\mu\text{l}$ . The ATPase activity was determined using the Baginski method described above in the presence of a final concentration of 40  $\mu\text{M}$   $\text{ZnCl}_2$  or 1 mM EDTA for background correction. One experiment with three replicates was performed for each of the ions tested ( $\text{Na}^+$  and  $\text{K}^+$ ).

**Effect of  $\text{Mg}^{2+}$  on *S. sonnei* ZntA activity in proteoliposomes.** The buffer was exchanged by diluting proteoliposome stocks in 20 mM MOPS, pH 6.8, 250 mM NaCl, 80 mM KCl, 5 mM  $\text{MgCl}_2$  and 1 mM DTT (Buffer MCA) or 20 mM MOPS, pH 6.8, 250 mM NaCl, 80 mM KCl and 1 mM DTT (Buffer MCB) followed by three freeze–thaw cycles and extrusion through 0.2- $\mu\text{m}$  polycarbonate filters. Proteoliposomes were collected by ultracentrifugation at 163,000g for 60 min at 4 °C and suspended in the corresponding buffer. The ATPase activity was determined using the Baginski method as described above. As the buffer used in the assays contains ATP and  $\text{MgCl}_2$  (required for ATP hydrolysis), the ATPase activity is stimulated exclusively for correctly oriented *S. sonnei* ZntA (N-domain facing outside). The presence or absence of  $\text{Mg}^{2+}$  in the proteoliposome lumen (buffer MCA or MCB) allows the identification of the putative requirement of  $\text{Mg}^{2+}$  counter-ion transport for activity. One experiment with three replicates was performed.

**Determination of  $\text{Zn}^{2+}$ -binding stoichiometry using Zincon.** *S. sonnei* ZntA and ZntA mutants were titrated with 5–6  $\text{ZnCl}_2$  equivalents per mol (using a 10 mM  $\text{ZnCl}_2$  stock) and subsequently desalted in 20 mM MOPS-KOH, pH 6.8, 80 mM KCl, 100 mM NaCl, 3 mM  $\text{MgCl}_2$ , 0.15  $\text{mg ml}^{-1}$   $\text{C}_{12}\text{E}_8$  and 1 mM TCEP using a HiTrap desalting column packed with Sephadex G-25 resin to remove free or loosely bound metal. The  $\text{Zn}^{2+}$  content of the samples was determined by colorimetric quantification upon complex formation with 2-[5-(2-hydroxy-5-sulphophenyl)-3-phenyl-1-formazyl]benzoic acid (Zincon). Briefly, metal release was achieved upon incubation in a final concentration of 30 mM HCl. Subsequently, samples were diluted to a final concentration of 100 mM borate, pH 9, and 4 M guanidinium chloride, followed by the addition of Zincon to a final concentration of 40  $\mu\text{M}$ . The quantification of Zincon– $\text{Zn}^{2+}$  complexes was performed in a 96-well plate reader (Perkin Elmer) by measurement of the absorbance at 630 nm using a calibration curve obtained by the addition of an increasing amount of  $\text{ZnCl}_2$  in the same buffer. The protein concentration was determined by a modified Bradford assay. Protein solutions (10  $\mu\text{l}$ ) were incubated with 10  $\mu\text{l}$  1 M NaOH. Subsequently, 500  $\mu\text{l}$  Bradford reagent was added, and quantification was performed in a 96-well plate reader by measurement of the absorbance at 600 nm using a calibration curve obtained with BSA standards. Three independent experiments with three replicates for each experiment were conducted.

**ClusPro docking.** *S. sonnei* ZntA in the E2P state and the *E. coli* ZntA HMBD fragment containing residues 46–118 (PDB ID, 1MWY<sup>13</sup>) were chosen. The sequence identity of the *E. coli* ZntA HMBD with the corresponding residues of the *S. sonnei* ZntA HMBD was 97%. Docking was done using the ClusPro server<sup>36</sup>. The best model in the van der Waals + electrostatics scoring scheme was selected, as judged by cluster size scores.

**Molecular dynamics simulations.** Two 60-ns atomistic molecular dynamics simulations were run, one for each of the two ZntA structures.  $\text{AlF}_4^-$  bound to the E2P structure was modelled as  $\text{H}_2\text{PO}_4^-$  as described previously<sup>37</sup>. D436–BeF<sub>3</sub><sup>−</sup> in the E2P structure was modelled as a phosphorylated aspartate using CHARMM27 parameters<sup>38</sup>. The bound  $\text{Mg}^{2+}$  was retained in both structures. ZntA was embedded in a dioleoylphosphatidylcholine (DOPC) membrane based on the coordinates of a

pre-equilibrated slab multiplied eight times from the Laboratory of Molecular & Thermodynamic Modelling, and the proteins were positioned according to transmembrane alignment with the Orientations of Proteins in Membranes database coordinates of the E2P-P<sub>i</sub> Cu<sup>+</sup>-ATPase structure (PDB ID, 3RFU<sup>11</sup>)<sup>39</sup>. Lipids with atoms within 0.8 Å of any protein atom were deleted. Finally, the protein–membrane systems were further solvated with TIP3P<sup>40</sup> water and neutralized with sodium.

Molecular dynamics simulations were run using the NAMD 2.8 program<sup>41</sup> employing the CHARMM27 force field for proteins<sup>42</sup> and the CHARMM36 force field for lipids<sup>43</sup>. Before simulation, the systems were subjected to 2,000 steps of conjugate gradient minimization. Then, a 0.5-ns molecular dynamics simulation was performed where everything but the lipid tails was kept constant (NVT ensemble,  $T = 298$  K), allowing the lipids to adapt to the protein to some extent. Next, the systems were minimized for 1,000 steps after which all atoms were allowed to move freely for 0.5 ns (NPT ensemble,  $T = 298$  K,  $P = 1$  atm) except for the protein backbones, which were held fixed. Finally, all atoms were allowed to move freely in a production run of 60 ns. The temperature was controlled by Langevin dynamics, and the Nosé–Hoover–Langevin piston method was used for controlling the pressure<sup>44,45</sup>. The electrostatics were fully accounted for by applying the particle mesh Ewald method with periodic boundary conditions<sup>46</sup>. The van der Waals interactions were truncated at 12 Å, applying a switching function at 10 Å. The neighbour list containing all pairs of atoms for which non-bonded interactions are calculated included atoms within 14 Å of each other and was updated for every 20 steps. Bonded interactions were evaluated every 1 fs, while electrostatic and van der Waals interactions were evaluated every 4 and 2 fs, respectively. Each production run was for 60 ns, producing 60,000 frames, of which 2,000, evenly spread over the simulation time, were used for analysis.

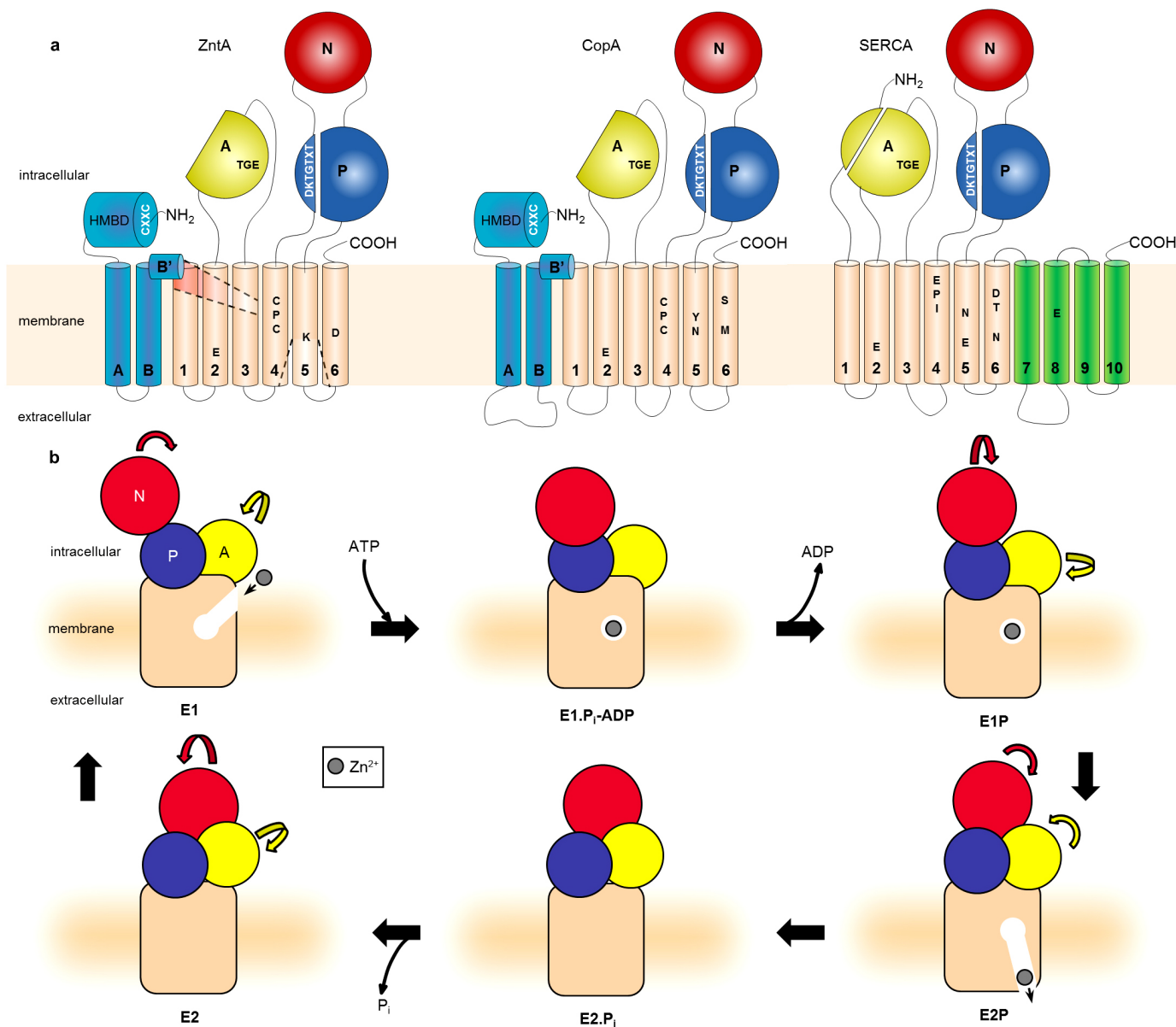
To describe the release pathway and accompanying  $\text{Zn}^{2+}$ –protein interactions, a steered molecular dynamics (SMD) approach was used<sup>47,48</sup>. A divalent  $\text{Zn}^{2+}$  was placed between the ion-coordinating residues C392, C394 and D714 in the E2P state, and this was followed by deletion of three clashing water molecules and a 10,000-step conjugate-gradient energy minimization. A force constant of 5 kcal mol<sup>−1</sup> Å<sup>−2</sup> and velocities of 10–20 Å ns<sup>−1</sup> were applied to the ion directed from the inside out in the  $z$ -direction in ten independent 1-ns simulations. Similar release pathways were observed in the ten SMD simulations, and the number of  $\text{Zn}^{2+}$ –protein interactions within a 5 Å cut-off were calculated. The C $\alpha$ s of six remote residues (148, 198, 367, 383, 385 and 705) were restrained to keep the system from drifting when applying the force.

**Figures.** Structural representations were generated using PyMol<sup>49</sup>. Helices MA, MB and MB' have been removed for clarity in Fig. 1b, c, and helices M3–M4 of *L. pneumophila* CopA and *S. sonnei* ZntA were aligned to generate Fig. 1c. In Fig. 3c, the structures were aligned using the M4–M6 transmembrane helices, and the view is from the extracellular side. Taking the E1–E2 conformational changes into account, K693–D714 (*S. sonnei* ZntA) and R655–D684 (*A. thaliana* AHA2) almost overlap. In H<sup>+</sup>-ATPases, *A. thaliana* AHA2 D684 participates in H<sup>+</sup> transfer to the extracellular side, and R655 has been proposed to stimulate H<sup>+</sup> release from D684 and prevent re-protonation<sup>18</sup>. F210 of *S. sonnei* ZntA separates the electronegative ion entry funnel from the membranous  $\text{Zn}^{2+}$  site and overlaps with *A. thaliana* AHA2 N106, which stabilizes the protonated AHA2 D684 (ref. 18) and blocks intracellular H<sup>+</sup> exchange<sup>24</sup>. In Fig. 3d, the electron density is provided for the E2P state. A deep pathway reaches the intramembranous high-affinity ion-binding site and may allow  $\text{Zn}^{2+}$  release via E202. In Fig. 3e, the view is from the intracellular side. Ion entry to *S. sonnei* ZntA may occur through negatively charged residues placed inside the periphery of the positively charged residues of MB'. The view in Fig. 3f is identical to that in Fig. 3e, displaying a highly electronegative funnel (negative surface in red and positive in blue) formed by the residues of M1, M2 and M3. The funnel extends towards the ion-binding CPC motif and is constricted by M187 and F210, presumably guiding ions to the membranous high-affinity ion-binding site and excluding non-transported compounds (see also Extended Data Table 2b). The view in Fig. 3g is identical to that in Fig. 3e but for the Cu<sup>+</sup>-ATPase CopA of *L. pneumophila* (PDB ID, 3RFU<sup>11</sup>), for which ion uptake has been proposed to occur instead through a transient Cu<sup>+</sup>-binding site.

- Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).
- McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
- Afonine, P. V. et al. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D* **68**, 352–367 (2012).
- Baginski, E. S., Foa, P. P. & Zak, B. Microdetermination of inorganic phosphate, phospholipids, and total phosphate in biologic materials. *Clin. Chem.* **13**, 326–332 (1967).
- Kozakov, D. et al. Achieving reliability and high accuracy in automated protein docking: ClusPro, PIPER, SOU, and stability analysis in CAPRI rounds 13–19. *Proteins* **78**, 3124–3130 (2010).



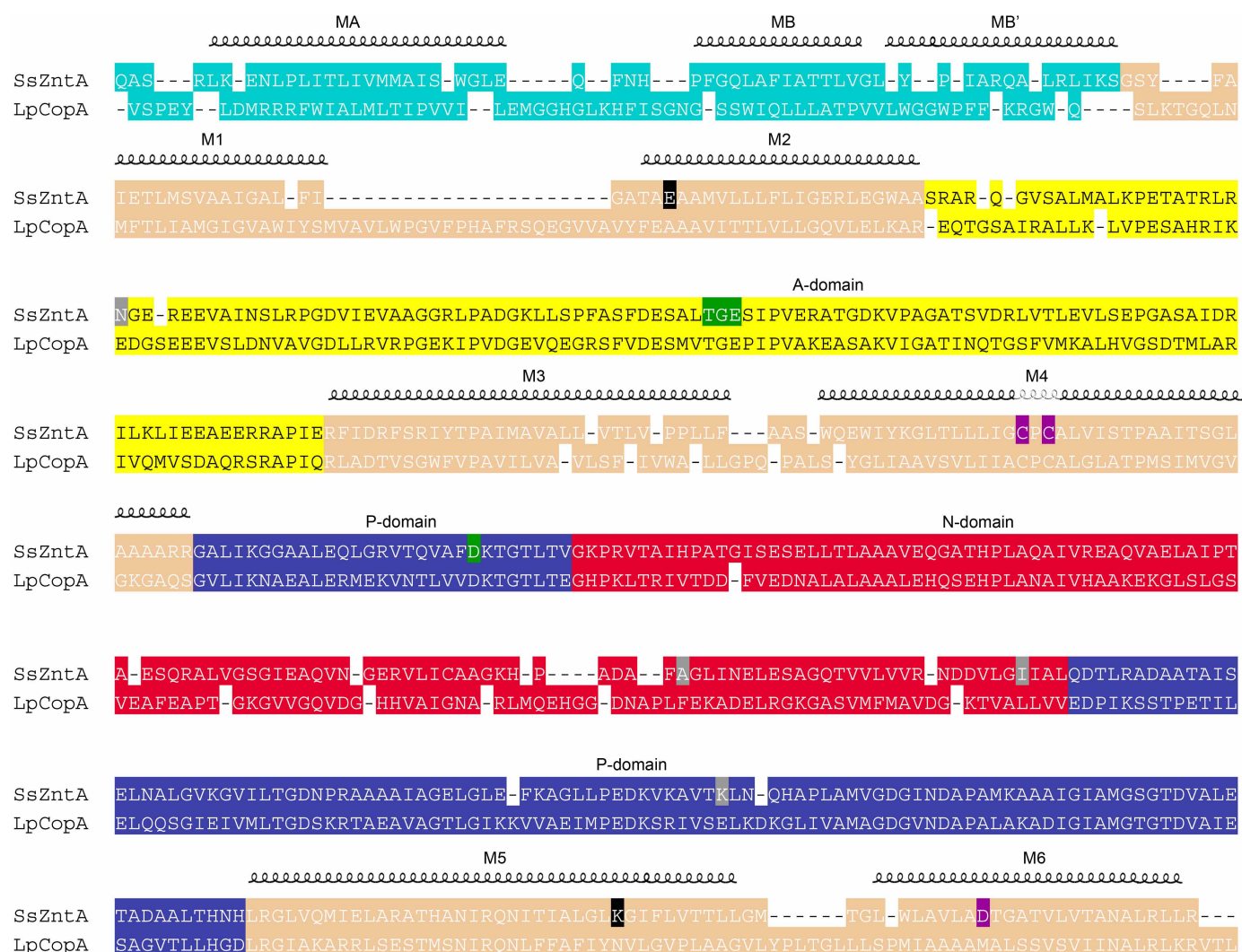
37. Yang, W., Gao, Y. Q., Cui, Q., Ma, J. & Karplus, M. The missing link between thermodynamics and structure in F1-ATPase. *Proc. Natl Acad. Sci. USA* **100**, 874–879 (2003).
38. Damjanović, A., Garcia-Moreno, E. B. & Brooks, B. R. Self-guided Langevin dynamics study of regulatory interactions in NtrC. *Proteins* **76**, 1007–1019 (2009).
39. Lomize, M. A., Lomize, A. L., Pogozheva, I. D. & Mosberg, H. I. OPM: orientations of proteins in membranes database. *Bioinformatics* **22**, 623–625 (2006).
40. Jorgensen, W., Chandrasekhar, J., Madura, J., Impey, R. & Klein, M. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
41. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).
42. MacKerell, A. D. *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586–3616 (1998).
43. Klauda, J. B. *et al.* Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *J. Phys. Chem. B* **114**, 7830–7843 (2010).
44. Martyna, G. J., Tobias, D. J. & Klein, M. L. Constant-pressure molecular-dynamics algorithms. *J. Chem. Phys.* **101**, 4177–4189 (1994).
45. Feller, S. E., Zhang, Y. H., Pastor, R. W. & Brooks, B. R. Constant-pressure molecular-dynamics simulation: the Langevin piston method. *J. Chem. Phys.* **103**, 4613–4621 (1995).
46. Essmann, U. *et al.* a smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593 (1995).
47. Sotomayor, M. & Schulten, K. Single-molecule experiments *in vitro* and *in silico*. *Science* **316**, 1144–1148 (2007).
48. Izrailev, S., Stepaniants, S., Balsera, M., Oono, Y. & Schulten, K. Molecular dynamics study of unbinding of the avidin–biotin complex. *Biophys. J.* **72**, 1568–1581 (1997).
49. The PyMOL molecular graphics system v.1.3r1 (Schrödinger, LLC, 2010).
50. Wu, C. C. *et al.* The cadmium transport sites of CadA, the Cd<sup>2+</sup>-ATPase from *Listeria monocytogenes*. *J. Biol. Chem.* **281**, 29533–29541 (2006).
51. Post, R. L., Kume, S. & Hegyvary, C. Activation by adenosine triphosphate in phosphorylation kinetics of sodium and potassium ion transport adenosine triphosphatase. *J. Biol. Chem.* **247**, 6530–6540 (1972).
52. Braberg, H. *et al.* SALIGN: a web server for alignment of multiple protein sequences and structures. *Bioinformatics* **28**, 2072–2073 (2012).
53. Sharma, R., Rensing, C., Rosen, B. P. & Mitra, B. The ATP hydrolytic activity of purified ZntA, a Pb(II)/Cd(II)/Zn(II)-translocating ATPase from *Escherichia coli*. *J. Biol. Chem.* **275**, 3873–3878 (2000).
54. von Heijne, G. The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology. *EMBO J.* **5**, 3021–3027 (1986).



### Extended Data Figure 1 | Topology and reaction cycle of P-type ATPases.

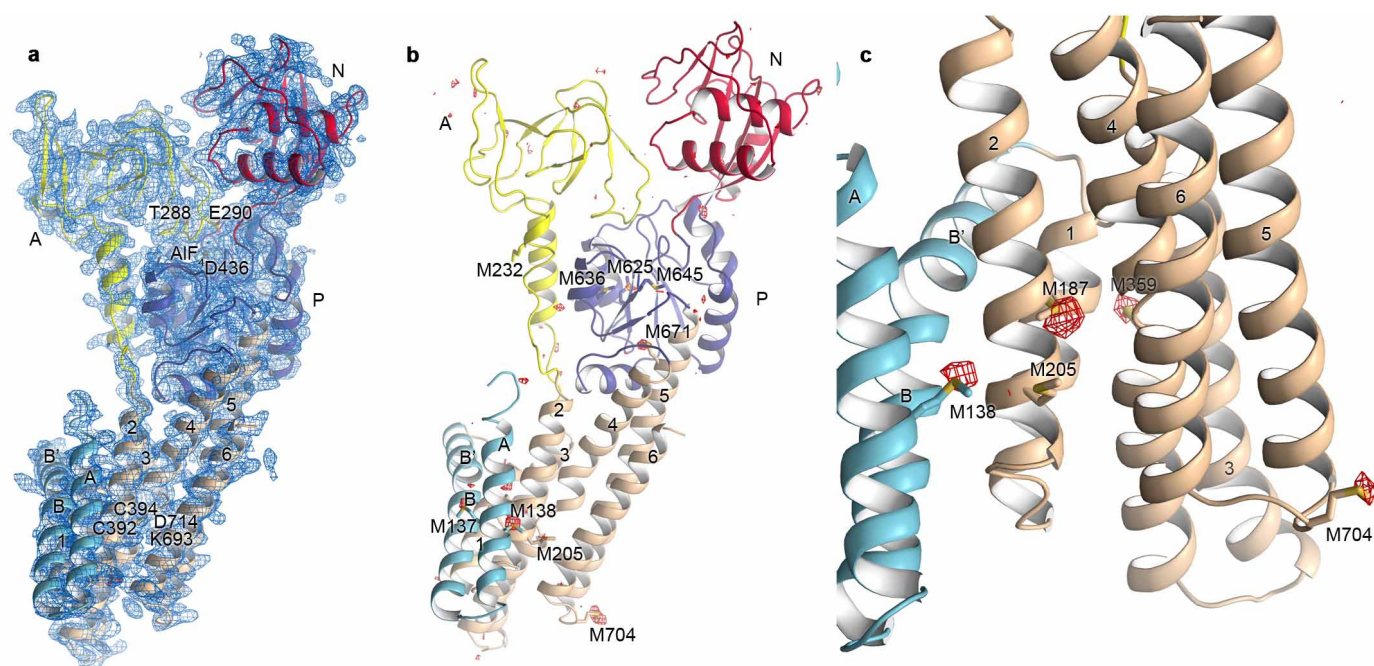
**a**, Topology of ZntA, CopA and SERCA. Key residues in the HMBD and A, P, N and M domains are highlighted. In ZntA, the negatively charged ion entry funnel and release pathway are outlined. D436 in *S. sonnei* ZntA is the autophosphorylated/dephosphorylated catalytic aspartate in the DKTGTXI motif of the P domain. C392 and C394 in M4, K693 in M5 and D714 in M6 of *S. sonnei* ZntA have been proposed to bind zinc in biochemical studies<sup>10,15,50</sup>. **b**, The Post-Albers (E1 to E2) reaction cycle of Zn<sup>2+</sup>-transporting P-type ATPase<sup>9,51</sup>. Phosphorylation events in the intracellular domains drive large

conformational changes that permit alternating access to transport sites in the membrane about 50 Å from the ATP-targeted catalytic aspartate. According to the model, a high-affinity state (E1), which is open to the intracellular space, binds to Zn<sup>2+</sup> and enters an occluded state. This state then undergoes phosphorylation. Completion of this event (E1P) triggers the release of the Zn<sup>2+</sup>, establishing an outward-facing, low-affinity state (E2P). Release of the inorganic phosphate (P<sub>i</sub>) yields the fully dephosphorylated conformation (E2), which is followed by restoration of the inward-facing conformation (E1), which initiates a new reaction cycle.



**Extended Data Figure 2 | Structure-based sequence alignment of *S. sonnei* ZntA and *L. pneumophila* CopA.** Helix positions are indicated for *S. sonnei* ZntA, and noteworthy residues are highlighted. Four of seven amino acid positions in which ZntA differs between *S. sonnei* and *Escherichia coli* and that are likely to be functionally irrelevant are highlighted in grey. The high-affinity

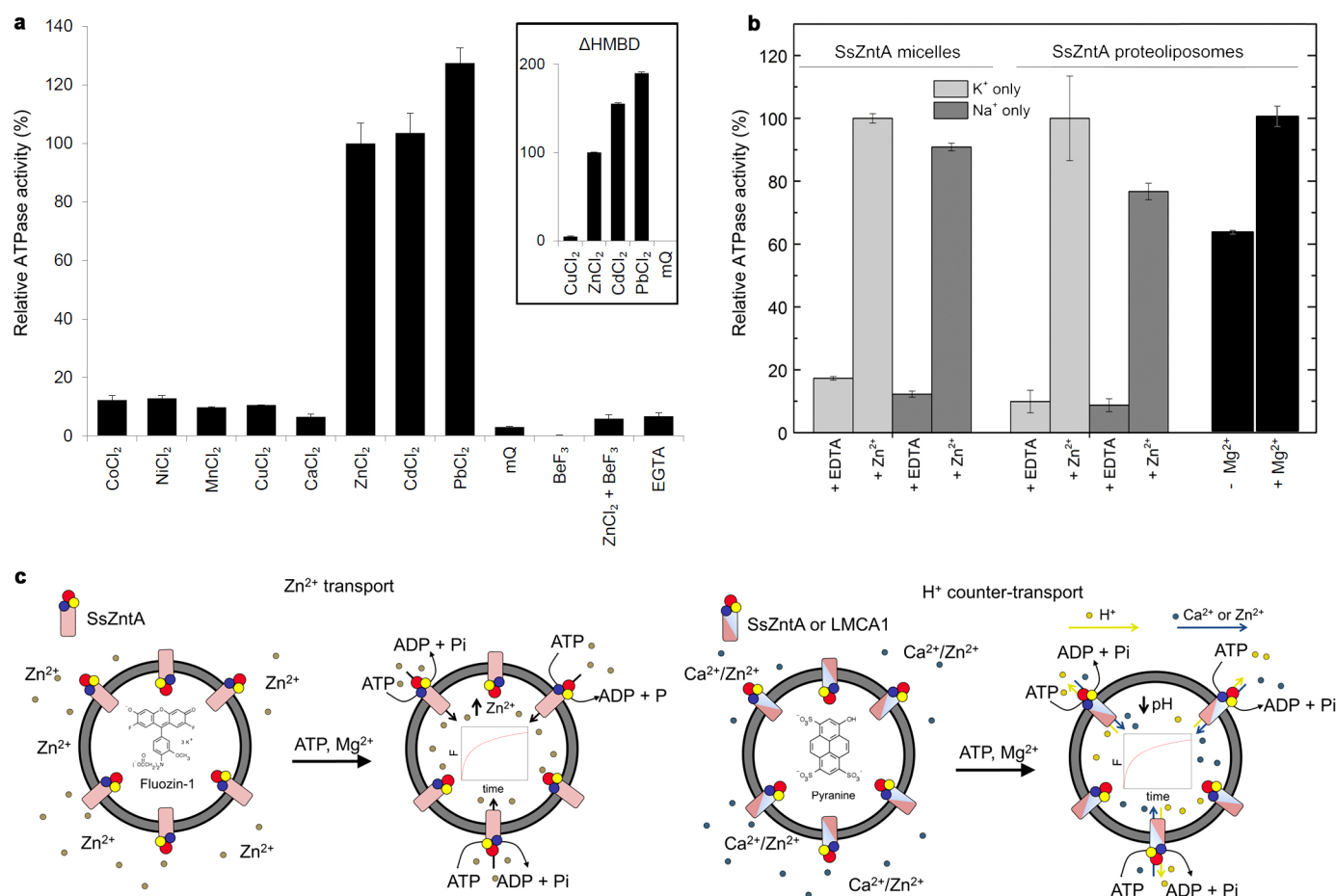
ion-binding residues C392, C394 and D714 are indicated in purple; the catalytically phosphorylated aspartate and the dephosphorylating TGE motif are highlighted in green. E202 and K693, which are possibly involved in ion release, are marked in black. The alignment was performed using SALIGN<sup>52</sup>.



**Extended Data Figure 3 | Electron density of the determined E2·P<sub>i</sub> state of *S. sonnei* ZntA.** **a**, Final  $2F_o - F_c$  electron density of *S. sonnei* ZntA in the E2·P<sub>i</sub> state. The density is contoured at  $1\sigma$ , and the view is equivalent to that shown in Fig. 1a. **b**, **c**, Se-Met peaks calculated using Se-SAD (Se

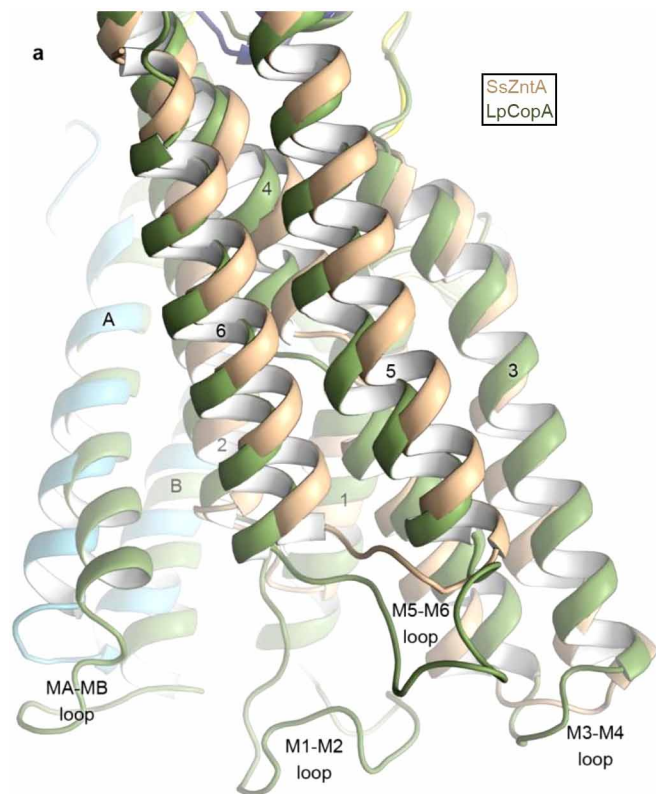
single-wavelength anomalous diffraction) data and phases obtained from molecular-replacement-guided model building. The anomalous difference Fourier map is contoured at  $3\sigma$ . A view of the entire protein (**b**), and a view of the M domain (**c**) are shown.





**Extended Data Figure 4 | Functional assays of *S. sonnei* ZntA.** **a**, Wild-type and ΔHMBD (inset) *S. sonnei* ZntA ATPase activity is stimulated by Zn<sup>2+</sup>, Cd<sup>2+</sup> and Pb<sup>2+</sup>. ATPase activity (normalized; the activity in the presence of Zn<sup>2+</sup> is set at 100% for the wild type and ΔHMBD, respectively) was determined using the Baginski assay (see Methods for details). This ion stimulation profile matches the one observed for ZntA from *E. coli*<sup>53</sup>. The mean + s.d. of technical replicates is shown ( $n = 3$ ). **b**, Effect of K<sup>+</sup>, Na<sup>+</sup> and Mg<sup>2+</sup> on *S. sonnei* ZntA activity. The ATPase activity of wild-type *S. sonnei* ZntA in detergent micelles or upon reconstitution in proteoliposomes, in

buffers containing exclusively Na<sup>+</sup> or K<sup>+</sup>, as determined by the Baginski assay, is shown. For Mg<sup>2+</sup>, the activity was in the proteoliposomes for internal buffers with or without MgCl<sub>2</sub>. The mean + s.d. of technical replicates is shown ( $n = 3$ ). **c**, Zn<sup>2+</sup> and H<sup>+</sup> transport across vesicle membranes. Zn<sup>2+</sup> transport of wild-type and D436N *S. sonnei* ZntA proteoliposomes monitored using the zinc-selective chelator FluoZin-1 (left). H<sup>+</sup> counter-ion transport in wild-type *S. sonnei* ZntA or Ca<sup>2+</sup>-ATPase LMCA proteoliposomes monitored using the pH indicator pyranine (right).



**b**

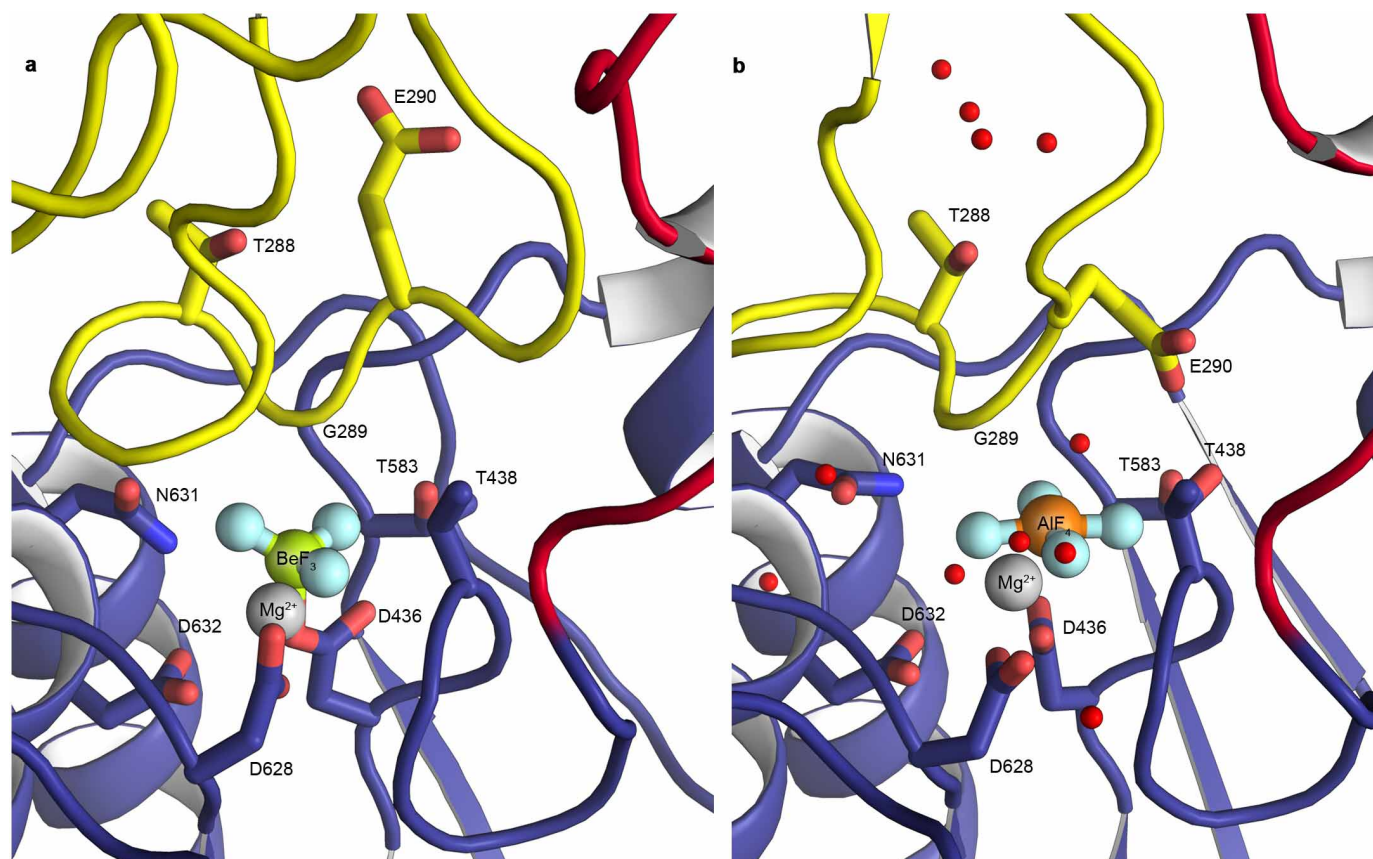
Loop	SsZntA		
	Length	On average in ZntA	Standard deviation
MA-MB	5	5.2	5.1
M1-M2	6	6.0	0.5
M3-M4	7	7.4	2.7
M5-M6	8	8.0	0.0

**c**

Loop	LpCopA		
	Length	On average in CopA	Standard deviation
MA-MB	11	20.1	6.0
M1-M2	19	13.3	5.4
M3-M4	7	13.4	8.1
M5-M6	14	11.9	4.8

**Extended Data Figure 5 | Structural comparison of ZntA and CopA.**  
**a**, Difference between the extracellular loops of *S. sonnei* ZntA and *L. pneumophila* CopA. *S. sonnei* ZntA is coloured as in Fig. 1a and *L. pneumophila* CopA is in dark green, and the proteins have been aligned on helices M5 and M6. Note that the loops are substantially longer in *L. pneumophila* CopA than in *S. sonnei* ZntA, which is a conserved difference

between Cu<sup>+</sup>- and Zn<sup>2+</sup>-transporting P-type ATPases (see also **b**).  
**b, c**, Comparison of the extracellular loop lengths of ZntA (**b**) and CopA (**c**). The lengths of the loops in *S. sonnei* ZntA and *L. pneumophila* CopA are shown, as well as averages based on 521 ZntA-type proteins and 617 CopA-type proteins (with less than 99% and 95% sequence identity within the ZntA and CopA sequences, respectively).

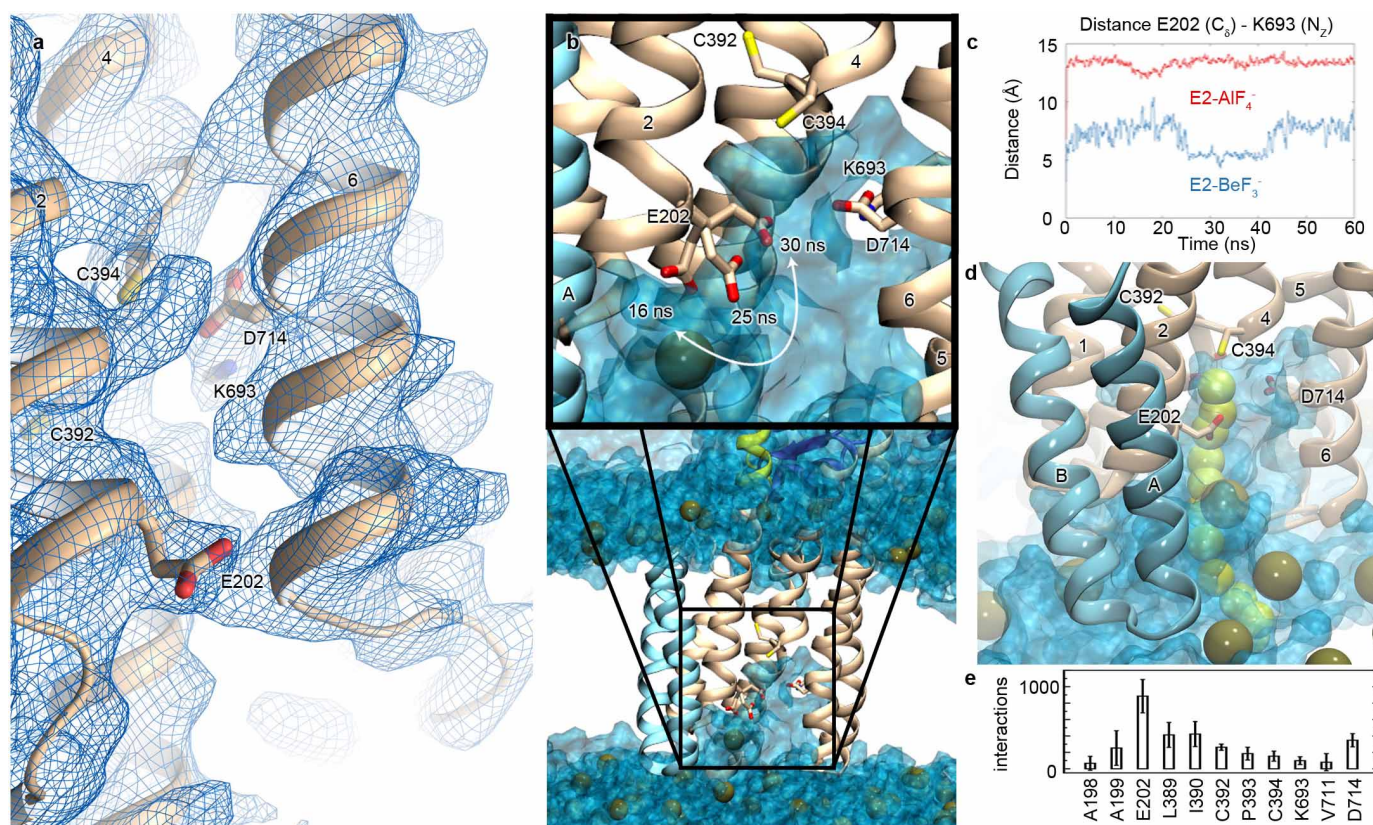


#### Extended Data Figure 6 | The phosphorylation site of *S. sonnei* ZntA.

The domains are coloured as in Fig. 1a.  $\text{AlF}_4^-/\text{BeF}_3^-$  (Al in orange, Be in green and F in cyan) and the  $\text{Mg}^{2+}$  ion (grey) are associated with D436 (in the DKTGTXT motif of the P domain) at the interface between the A and P domains. D436, T438, T583, D628, N631 and D632 (in the P domain), as well as T288, G289 and E290 (the TGE motif in the P domain that is associated with

dephosphorylation), are shown as sticks. Water molecules are shown as red spheres (not modelled for the E2P state). **a**, The E2P- $\text{BeF}_3^-$ -bound state. The catalytic D436 is protected from the TGE loop. **b**, The E2· $\text{P}_i$ - $\text{AlF}_4^-$ -bound state. E290 of the TGE loop probably activates a water molecule for dephosphorylation as observed in the equivalent E2· $\text{P}_i$  state of SERCA1A and CopA<sup>4,5,11</sup>.

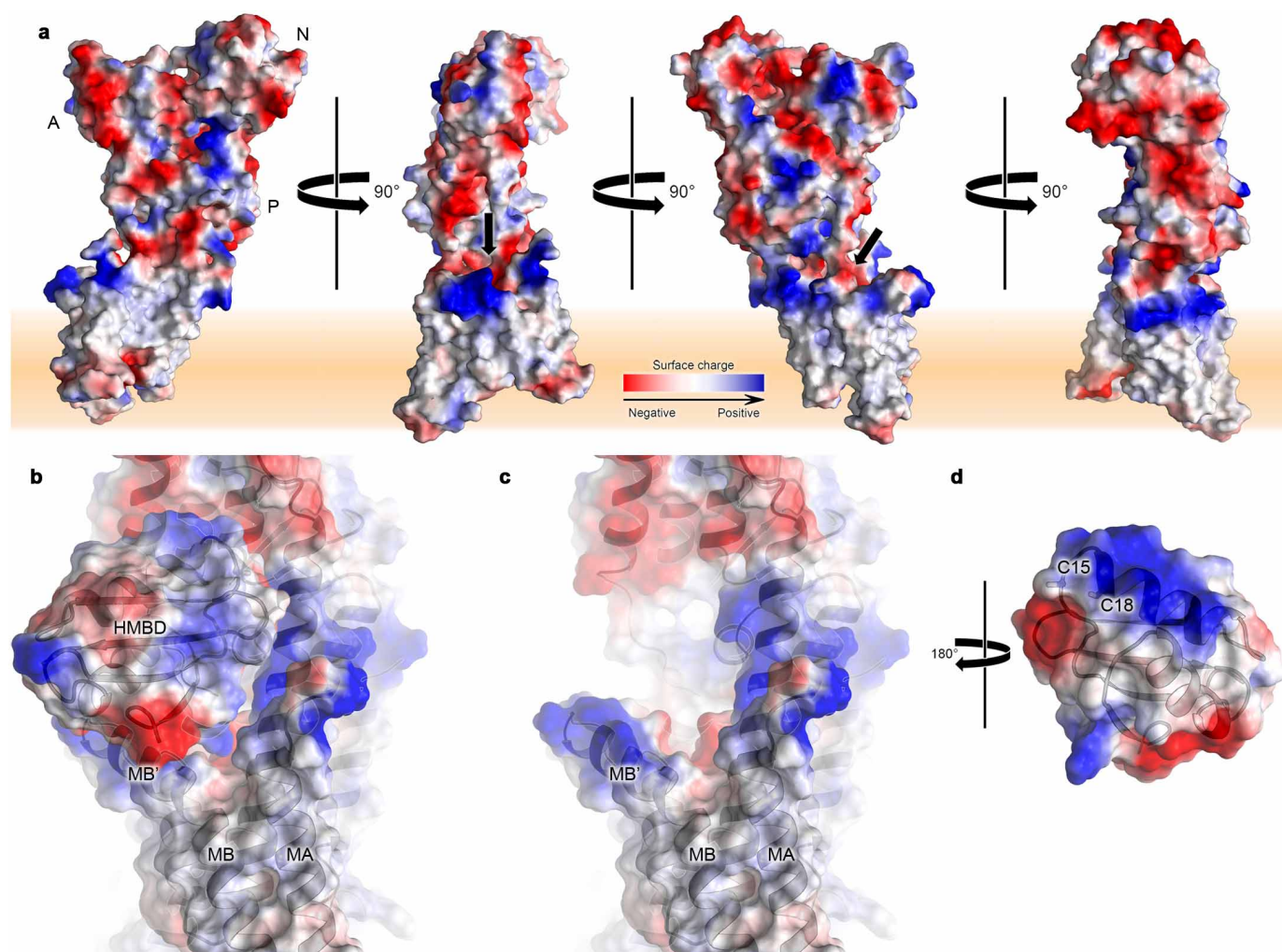




**Extended Data Figure 7 | The extracellular pathway.** **a**, The extracellular fraction of the E2-AlF<sub>4</sub><sup>-</sup> crystal structure. Functionally important residues are shown as sticks, and the protein is coloured as in Fig. 1a. The final 2F<sub>o</sub> - F<sub>c</sub> electron density is contoured at 1σ. The view is equivalent to the one in Fig. 3d. **b**, Dynamics of E202 in a 60-ns molecular dynamics simulation of the E2-BeF<sub>3</sub><sup>-</sup> structure in a dioleoylphosphatidylcholine (DOPC) membrane in the absence of zinc. Selected residues are shown as sticks. Representative E202 conformations were captured at 16, 25 and 30 ns from snapshots aligned according to backbone Cαs of M1-M4. The orientation of E202 at 16 ns resembles how this side chain appears in the E2-AlF<sub>4</sub><sup>-</sup> state, while the flexibility observed throughout the simulation agrees with the observed poor electron density of the side chain in the E2-BeF<sub>3</sub><sup>-</sup> state (see Fig. 3b). Note that there are two distorted lipids at the release pathway that may assist in Zn<sup>2+</sup>

release (vdW spheres represent lipid phosphates). **c**, Distance between the centre of mass of the Cδ of the E202 side chain and the N<sub>ζ</sub> of the K693 side chain during the 60-ns simulations of the E2-AlF<sub>4</sub><sup>-</sup> and E2-BeF<sub>3</sub><sup>-</sup> *S. sonnei* ZntA structures in the absence of zinc, as a running average over five consecutive frames of each trajectory. **d**, The release pathway and accompanying protein interactions experienced by Zn<sup>2+</sup> in a steered molecular dynamics simulation originating from the centre of mass of residues C392, C394 and D714. The transmembrane domain, lipid phosphates and water within 7 Å of the protein are coloured as in **b**. **e**, The number of Zn<sup>2+</sup>-protein interactions with a 5 Å cut-off during steered molecular dynamics (SMD) simulations. Error bars correspond to counts from ten independent simulations with pulling speeds on Zn<sup>2+</sup> of 10–20 Å ns<sup>-1</sup>.





**Extended Data Figure 8 | Surface charge distribution and docking of the HMBD to *S. sonnei* ZntA.** **a**, Four views of the overall structure of E2-AlF<sub>4</sub><sup>-</sup>. The view to the left is equivalent to that in Fig. 1a. The charge distribution complies with the positive-inside rule for membrane proteins<sup>54</sup>. The putative ion entry funnel is indicated with a black arrow. **b–d**, Docking of the HMBD to *S. sonnei* ZntA. The apo-HMBD of *E. coli* ZntA (PDB ID, 1MWY<sup>13</sup>) docks to

the entry site region of *S. sonnei* ZntA using electrostatic complementation and van der Waals interactions, as predicted by the ClusPro 2.0 server<sup>36</sup> (**b**). Equivalent view to that in **a** of *S. sonnei* ZntA without the HMBD (**c**). View of the isolated HMBD, rotated 180° relative to **a** to show the surface complementary to *S. sonnei* ZntA (**d**). The ion-binding cysteine residues C15 and C18 are highlighted.

Extended Data Table 1 | Data collection, phasing and refinement statistics

Data collection*									
	E2-P <sub>i</sub> [AlF <sub>4</sub> <sup>-</sup> ]			E2P[BeF <sub>3</sub> <sup>-</sup> ]			Se-E2-P <sub>i</sub> [AlF <sub>4</sub> <sup>-</sup> ]		
	C222 <sub>1</sub>			P2 <sub>1</sub>			C222 <sub>1</sub>		
Space group									
Cell dimensions									
<i>a</i> , <i>b</i> , <i>c</i> (Å)	77.6	83.6	319.8	54.5	61.0	141.5	76.1	82.2	314.8
<i>α</i> , <i>β</i> , <i>γ</i> (°)	90.0	90.0	90.0	90.0	96.0	90.0	90.0	90.0	90.0
Resolution, (Å)*	50-2.70 (2.80-2.70)			50-3.20 (3.30-3.20)			50-4.50 (4.62-4.50)		
<i>R</i> <sub>merge</sub> (%)	6.5 (126.7)			20.7 (115.3)			32.5 (99.1)		
<i>I</i> / <i>σI</i>	15.5 (1.04)			8.07 (1.37)			8.74 (3.40)		
Completeness (%)	99.7 (93.2)			99.7 (99.8)			99.8 (99.0)		
Redundancy	4.6 (4.6)			4.6 (4.8)			7.7 (7.8)		
CC(1/2) <sup>†</sup> (%)	99.9 (73.5)			99.0 (56.8)			99.1 (84.3)		
Refinement									
Resolution (Å)	50-2.70 (2.75-2.70)			50-3.20 (3.40-3.20)					
No. reflections	28862 (1294)			15448 (2437)					
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub> (%)	20.7/24.0 (25.9/48.7)			21.0/28.1 (26.9/33.9)					
No. atoms									
Protein	4448			4347					
Ligand/ion	6			5					
Water	56			0					
B-factor									
Protein	77.9			92.4					
Ligand/ion	46.3			76.4					
Water	55.5								
R.m.s deviations									
Bond lengths (Å)	0.005			0.004					
Bond angles (°)	0.932			0.841					

\* The highest resolution shell is shown in parenthesis. †CC<sub>1/2</sub> values were calculated using the program XDS.

Extended Data Table 2 | Statistical analysis of the ion entry region of *S. sonnei* ZntA

a

Block	Number of Asp + Glu in total	Number of sequences	Amino acid position	Number of Asp or Glu	Amino acid in SsZntA	
M1	0	6	182	264	A	
	1	357	183	227	I	
	2	158	184	144	E	
	3	0				
M2	0	35	210	0	F	
	1	472	211	14	L	
	2	14	214	485	E	
	3	0	215	1	R	
	4	0				
M3	0	234	345	4	R	
	1	287	347	0	I	
	2	0	348	283	D	
	3	0				
All	0	0	b	Number of Lys + Arg in MB'		Number of sequences
	1	8		0	0	
	2	155		1	18	
	3	328		2	130	
	4	30		3	308	
5	0	4		58		
		5		6		
		6		1		

c

Block	Number of Asp + Glu in total	Number of sequences	Amino acid position	Number of Asp or Glu	Amino acid in LpCopA
M1	0	617	153	0	A
	1	0	154	0	M
	2	0	155	0	G
	3	0			
M2	0	608	198	0	V
	1	8	199	1	L
	2	0	202	7	Q
	3	0	203	0	V
	4	0			
M3	0	31	334	25	R
	1	560	336	1	A
	2	26	337	586	D
	3	0			
All	0	31			
	1	552			
	2	34			
	3	0			
	4	0			
	5	0			

**a**, Conservation of the electronegative ion entry funnel in ZntA. The negative charges are provided in three blocks of surface-exposed residues in helices M1 (182, 183, 184), M2 (210, 211, 214, 215) and M3 (345, 347, 348), in the vicinity of the negatively charged entry funnel of *S. sonnei* ZntA. Five hundred and twenty-one ZntA-type proteins with less than 99% sequence identity, selected from the latest UniProt database, were used for the analysis. **b**, The number of positively charged residues in the MB' helix of ZntA proteins using the same data set as in **a**. **c**, Conservation of the CopA region equivalent to the electronegative ion entry funnel in ZntA. The number of negatively charged residues in the MB' helix of CopA proteins. Six hundred and seventeen CopA-type proteins with less than 95% sequence identity, selected from the latest UniProt database, were used for the analysis. (See also Fig. 3e, f.)

# CAREERS

**TURNING POINT** Catalysis led chemist to work in polymers **p.525**

**@NATUREJOBS** Follow us on Twitter for the latest on jobs and careers [go.nature.com/e492gf](http://go.nature.com/e492gf)

**NATUREJOBS** For the latest career listings and advice [www.naturejobs.com](http://www.naturejobs.com)



## CAREER CHANGES

# Open for business

*A master's in business can offer scientists extra flexibility or whole new career paths.*

BY KAREN RAVN

In June 2011, Karen Havenstrite graduated from Stanford University in California with a PhD in chemical engineering. A few months later, she had invented a coating that makes contact lenses more comfortable and had co-founded a company to sell it.

She soon discovered that starting a business is tough. "In the process of taking science from the

lab into the marketplace, I realized how much I didn't know," she says. She is learning it all now. In 2015, she will graduate from Stanford again — this time with a master's in business administration (MBA). At about the same time, her company — Ocular Dynamics in Menlo Park, California — will begin to market her invention.

Data from the US National Center for Education Statistics in Washington DC suggest that Havenstrite's MBA will be one of about 194,000

advanced business degrees awarded next June in the United States alone. Scientists make up a minority of MBA enrolment, but the degree is something that researchers should consider: it could facilitate advancement in their existing careers, or open up prospects by helping them to turn an idea into a business plan and then into a profitable venture (see 'Start-ups').

Many working scientists feel that they have already racked up enough years — and debt — getting PhDs. Yet for some, 'B-school' might be just the ticket, and there are ways to mitigate what can be formidable costs. Some employers will pay the tuition fees and promote the researchers when they complete the degree. And weekend classes offer a way to keep full-time jobs while completing the course. Time and money can also be saved by choosing a one-year programme instead of the more-traditional two-year stint. Online MBAs are usually the least expensive choice.

## POTENTIAL REWARDS

The degree can offer returns on the investment, however. The Graduate Management Admission Council, a non-profit organization based in Reston, Virginia, has found consistently high percentages of alumni who report that their degrees have paid off in terms of income and job satisfaction. In a survey of nearly 21,000 alumni of 132 business schools around the world, all graduating between 1959 and 2013, most said that their education had been rewarding personally (94%), professionally (90%) and financially (77%).

There is no dearth of programmes to choose from — 13,000 worldwide, according to the Association to Advance Collegiate Schools of Business in Tampa, Florida. And for vetting, an abundance of organizations and publications rank programmes according to various criteria, including cost, peer and recruiter assessment, starting salary and bonus, employment rates and test scores. The location of the university and the length of the programme are likely to be crucial in making the choice.

The MBA degree was created in the United States in 1908, and US universities have tended to dominate both in enrolments and in rankings. But institutions such as the London Business School; INSEAD in France, Singapore and Abu Dhabi; and the IE Business School in Spain have also found their ways onto some elite lists. People wanting a one-year programme are likely to have better luck looking in Europe, because shorter programmes are more popular there than in the United States. ►



## START-UPS

*Entrée to entrepreneurship*

Some of the biggest companies in the world — Amazon, Apple, Google — are famous for being born in garages. As yet, they are unrivalled by any of the fledgling companies born in Stanford University's Startup Garage in California. But even though the short course has been running for only two years, several businesses it has spawned are thriving, and a couple founded by scientists should be under way soon. Enara Health in San Mateo, California, will launch early next year and uses mobile technology to deliver ancillary health care interactively with the goal of improving access and follow-up for obesity and related conditions. "Students begin the course with a need

and a user in mind," says coordinator Ryann Price. In teams of 2–4, students design products, make prototypes, create business models, test hypotheses and seek funding. "We encourage them to build and test simply and cheaply," says Price, who reports that students generally find the process more difficult than they thought.

This autumn, Stanford is partnering with Peking University in Beijing, China, to offer The Startup Garage: The China Version. One team in that course is trying to adapt 23andMe — a company in Mountain View, California, that provides customers with ancestry-related genetic reports — for the China market. **K.R.**

► A one-year programme is often the best option for people who want to accelerate their careers, rather than switch them, says Douglas Stayman, former associate dean of MBA programmes at Cornell University's Samuel Curtis Johnson Graduate School of Management in Ithaca, New York, which offers both one- and two-year programmes. Stayman is now overseeing the MBA programme at Cornell Tech, which in 2017 will move to a new campus in New York City.

The IE Business School is one of a few high-rated schools to offer online MBAs. The online degree has the same admission requirements as the full-time on-campus programme (online students tend to be a few years older and have a couple more years of work experience).

Earlier this year, the Graduate Management Admission Council surveyed more than 3,000 students in their final year of business school at 111 universities around the world and found that nearly 60% already had job offers. About one-quarter of the offers were in finance and accounting, one-fifth each in consulting and products and services (such as supply-chain management, or getting goods to market) and 15% in technology. Only 5% of the offers were in health care or pharmaceutical drug development — prime possibilities for scientists — but students seeking jobs in those areas were among the most likely to get offers. Many schools say that 80–90% of their students will be employed by three months after their graduation.

"An MBA is the only advanced degree I know of that expands your opportunities rather than shrinks them," says Dan Madden, senior manager for strategic planning at Zoetis, an animal-health company in Florham Park, New Jersey. Before earning his MBA, Madden was a chemist for Schering-Plough (now part of Merck).

Costs vary from university to university and from nation to nation. In 2013, the US business and technology news website *Business Insider*

compared the costs for the first year of an MBA programme at 11 US schools. Tuition alone ranged from about US\$53,000 to \$65,000, but total costs, including room and board, insurance, supplies and miscellaneous fees ranged from about \$81,000 to nearly \$100,000.

At the London Business School, tuition — including reading materials but no other expenses — was £64,200 (US\$102,000) for the 15–21 month programme that started in August. At INSEAD, tuition fees for the class graduating in 2015 were €62,500 (\$79,000), but mandatory health-insurance and administration fees add another €800, and living expenses are estimated at €22,300 in Singapore and at €23,600 in France.

Sometimes, an MBA will expand a scientist's career opportunities so much that the science aspect loses much of its attraction. That happened for Yasar Awan. Now a manager at Raytheon in Boston, Massachusetts, overseeing the flow of materials from circuit boards to wires in support of a US Navy programme, Awan is

a 2014 graduate of a five-year programme at Pennsylvania State University's Smeal College of Business in University Park in which students earn a bachelor of science — his was in biology — and an MBA (see 'Highlights'). Along the way, Awan discovered possibilities that he might never have seen. "I don't really want to go into science any more," he admits. "I fell in love with supply chain."

## VALUABLE SKILLS

Scientists who get MBAs usually continue to think of themselves as scientists at some level, at least. And they are generally more comfortable explaining how a new product or technology works — whether to fellow employees or to external clients — than are non-scientists, says Mark Pauly, a health-care-management researcher at the University of Pennsylvania's Wharton School in Philadelphia.

In fact, "there are certain roles where we really request a science background", says Beth Keeler, vice-president for global acquisition and career planning at Pfizer in New York. "It's not required, but many times it's preferred."

Also, scientists are generally more analytical than the average business-school student, says finance manager Ian McFetridge, also at Pfizer. "They're good at breaking down problems into smaller pieces, good at seeing how different factors are related." On the other hand, he adds, "a lot of people coming out of science are not as natural at communicating and presenting ideas. But when it does happen that a scientist can communicate well, then you have someone who can take a complicated subject and simplify it for the audience, and that's exceptional." Such pairing of communication and science is particularly useful in consulting work, when MBA graduates need to explain to their clients how their suggestions for boosting revenue or efficiency will work.

As a research scientist in New York City, Brandan Hillerich developed new biotechnologies. He loved his work. But filing for patents and coming up with commercial strategies to

## HIGHLIGHTS

*Some MBA programmes of interest to scientists*

Two US MBA programmes were designed especially with scientists in mind. The one-year programme at Cornell University's Samuel Curtis Johnson Graduate School of Management in Ithaca, New York, is intended for those who have already spent a good deal of time and money getting advanced degrees but now want to expand their expertise into business. Students in this programme miss out on the internships that two-year students participate in during the summer between their first and second years, but they have the extra bonus of an intensive summer term

before the regular school year begins.

The Smeal College of Business at Pennsylvania State University in University Park offers a small selective programme that allows students to earn both a bachelor of science and an MBA in five years. Students satisfy the requirements for their undergraduate degrees in the first three years, and during the next two years they take the regular MBA programme with other MBA students. (If students so choose, they can take six years to complete the programme.) **K.R.**

market those technologies presented a whole new challenge. “To me, this was even more exciting than the bench science,” he says. And he wanted more of it. So in May, he began a one-year MBA programme at Cornell.

Some find business more exciting than science just because things can happen faster. In 2004, Ana Albir graduated from the Massachusetts Institute of Technology in Cambridge with a major in physics. But the research experience she had accumulated as an undergraduate led her to decide that a career in physics was not for her. “It can take decades to see results,” she says. In 2009, she graduated from Stanford with an MBA and she is now chief executive of Moon-drop Entertainment, a company she founded in 2012 to create educational tablet apps for children. She still works on challenging problems, but now she can solve them in days or weeks simply by designing a clever piece of software. “With physics, I love the field,” she explains, “but in what I do now, I love the field and the pace.”

Business requires as much creativity as science, say many PhD graduates who are pursuing MBAs. But advanced science degrees tend to be more of an individual pursuit, whereas business qualifications usually involve working as part of a team. “I like the collaboration,” says Drew Rattigan, a second-year MBA candidate at Smeal. “I like a more social environment.”

Teamwork is a big plus for Hillerich too. After he had been in the programme at Cornell for less than a month, he knew all his classmates — at around 100, at least during the summer, a much larger number than in a PhD programme. Working in groups with them is great, he says. “Everyone thinks about things so differently. It expands your own thinking.”

An MBA was always on the cards for Ally Chang, who received a PhD in biomedical science from the University of Auckland in New Zealand in 2009 and an MBA from Cornell in 2011 and is now the new-products commercial manager at Corning Life Sciences in Tewksbury, Massachusetts. Her science background is a big asset when she works with researchers to decide whether their ideas will work in the marketplace. “I’ve heard the comment — so many PhDs, so few professorships,” Chang says. “But even before I started my PhD, I knew I didn’t want to be a professor. I wanted to do what I’m doing.” Chang has always had a passion for science — but she wants to turn her scientific ideas into commercial products.

Havenstrite and Hillerich, too, have chosen business for business’s sake, because they believe it is a way to have a direct, positive impact on people’s lives. “That is why I got into science in the first place,” Hillerich says. It is a sentiment expressed by many scientists who have, or are seeking, MBAs: they want to do work that has tangible, measurable effects, and soon, not in some abstract, distant future. ■

**Karen Ravn** is a freelance writer in Pacific Grove, California.

## TURNING POINT

# Andrew Dove

*Andrew Dove was named the 2014 Royal Society of Chemistry Gibson-Fawcett Award winner in May. A chemist at the University of Warwick, UK, Dove describes his circuitous path into research on biodegradable materials for regenerative medicine, which involves replacing or regenerating human tissue.*



### What area of chemistry first drew you in?

In a word, catalysis — designing inorganic catalysts that boost the efficiency or change the chemical properties of large polymers known as plastics. After working at BP Chemicals in Saltend, UK, during my fourth year of university, I thought I wanted to work on industry-sponsored projects — for example, using these catalysts to make polyethylene, a chemically resistant plastic.

### Why did you initially focus on industry?

It was probably my dad’s influence. Academia was not on my list of potential careers. But I came to realize that I really enjoyed basic research and wanted to give it a go. I applied for a PhD at Imperial College London, where my adviser offered me a project making polylactide, which is now the most widely used biopolymer around, particularly in biomedical applications. Now that it can be made from corn, rather than from petrochemicals, it is cheaper to use in the face of rising oil prices.

### How was your postdoc a turning point?

My wife and I moved to the United States to pursue postdoc positions in a bid to build up our CVs. I was at Stanford University in California for about 15 months working on inorganic catalysis. Then my funding ran out. But my wife still had her postdoc funding to work at IBM, and I was able to get a postdoc contract there too, in the company’s Center on Polymer Interfaces and Macromolecular Assemblies, which is funded by the US National Science Foundation. There, I started doing more organic catalysis. I had freedom to do whatever I wanted as long as good, publishable science was the result. It was a breakthrough period because it helped me to believe that I had good ideas and could translate them into interesting projects.

### Where did your research go from there?

I should credit the American Chemical Society with my change in direction. The inorganic chemistry and polymer talks in their meetings were always at opposite ends of

the conference centre when I attended them in 2003 and 2004 — so I had to choose which I found more interesting, and polymers won. Those meetings proved crucial for helping me to understand where the cutting edge for creating new polymer materials really was at the time. I saw a couple of opportunities. One was to find ways to give biodegradable materials different physical properties and use those materials in high-value applications such as in biomedical devices.

### How did you approach your job search?

I applied for jobs probably even before I was ready for them, and found that it really helped me to hone my research proposals. In 2004, I started applying for academic posts. Rather than looking for jobs with ‘inorganic’ in the advert, I applied for a UK fellowship in nanoscience, and persuaded the university that I had the skills for the job. In 2005, I started my own group at the University of Warwick.

### What are you working on at the moment?

I’m working on degradable polymers. One is a hydrogel material that could one day be combined with adult stem cells to make a scaffold able to regenerate a human spinal disc. Once the cells start to grow, the biological material would take over, leaving nothing synthetic in the body.

### How do you feel about media observations that Royal Society of Chemistry award winners often become Nobel laureates?

I find it quite amusing. It would be lovely if that happened, but I think the press made a bit of a leap. I don’t feel daunted by it because I don’t take it seriously. ■

INTERVIEW BY VIRGINIA GEWIN



market those technologies presented a whole new challenge. “To me, this was even more exciting than the bench science,” he says. And he wanted more of it. So in May, he began a one-year MBA programme at Cornell.

Some find business more exciting than science just because things can happen faster. In 2004, Ana Albir graduated from the Massachusetts Institute of Technology in Cambridge with a major in physics. But the research experience she had accumulated as an undergraduate led her to decide that a career in physics was not for her. “It can take decades to see results,” she says. In 2009, she graduated from Stanford with an MBA and she is now chief executive of Moon-drop Entertainment, a company she founded in 2012 to create educational tablet apps for children. She still works on challenging problems, but now she can solve them in days or weeks simply by designing a clever piece of software. “With physics, I love the field,” she explains, “but in what I do now, I love the field and the pace.”

Business requires as much creativity as science, say many PhD graduates who are pursuing MBAs. But advanced science degrees tend to be more of an individual pursuit, whereas business qualifications usually involve working as part of a team. “I like the collaboration,” says Drew Rattigan, a second-year MBA candidate at Smeal. “I like a more social environment.”

Teamwork is a big plus for Hillerich too. After he had been in the programme at Cornell for less than a month, he knew all his classmates — at around 100, at least during the summer, a much larger number than in a PhD programme. Working in groups with them is great, he says. “Everyone thinks about things so differently. It expands your own thinking.”

An MBA was always on the cards for Ally Chang, who received a PhD in biomedical science from the University of Auckland in New Zealand in 2009 and an MBA from Cornell in 2011 and is now the new-products commercial manager at Corning Life Sciences in Tewksbury, Massachusetts. Her science background is a big asset when she works with researchers to decide whether their ideas will work in the marketplace. “I’ve heard the comment — so many PhDs, so few professorships,” Chang says. “But even before I started my PhD, I knew I didn’t want to be a professor. I wanted to do what I’m doing.” Chang has always had a passion for science — but she wants to turn her scientific ideas into commercial products.

Havenstrite and Hillerich, too, have chosen business for business’s sake, because they believe it is a way to have a direct, positive impact on people’s lives. “That is why I got into science in the first place,” Hillerich says. It is a sentiment expressed by many scientists who have, or are seeking, MBAs: they want to do work that has tangible, measurable effects, and soon, not in some abstract, distant future. ■

**Karen Ravn** is a freelance writer in Pacific Grove, California.

## TURNING POINT

# Andrew Dove

*Andrew Dove was named the 2014 Royal Society of Chemistry Gibson-Fawcett Award winner in May. A chemist at the University of Warwick, UK, Dove describes his circuitous path into research on biodegradable materials for regenerative medicine, which involves replacing or regenerating human tissue.*



### What area of chemistry first drew you in?

In a word, catalysis — designing inorganic catalysts that boost the efficiency or change the chemical properties of large polymers known as plastics. After working at BP Chemicals in Saltend, UK, during my fourth year of university, I thought I wanted to work on industry-sponsored projects — for example, using these catalysts to make polyethylene, a chemically resistant plastic.

### Why did you initially focus on industry?

It was probably my dad’s influence. Academia was not on my list of potential careers. But I came to realize that I really enjoyed basic research and wanted to give it a go. I applied for a PhD at Imperial College London, where my adviser offered me a project making polylactide, which is now the most widely used biopolymer around, particularly in biomedical applications. Now that it can be made from corn, rather than from petrochemicals, it is cheaper to use in the face of rising oil prices.

### How was your postdoc a turning point?

My wife and I moved to the United States to pursue postdoc positions in a bid to build up our CVs. I was at Stanford University in California for about 15 months working on inorganic catalysis. Then my funding ran out. But my wife still had her postdoc funding to work at IBM, and I was able to get a postdoc contract there too, in the company’s Center on Polymer Interfaces and Macromolecular Assemblies, which is funded by the US National Science Foundation. There, I started doing more organic catalysis. I had freedom to do whatever I wanted as long as good, publishable science was the result. It was a breakthrough period because it helped me to believe that I had good ideas and could translate them into interesting projects.

### Where did your research go from there?

I should credit the American Chemical Society with my change in direction. The inorganic chemistry and polymer talks in their meetings were always at opposite ends of

the conference centre when I attended them in 2003 and 2004 — so I had to choose which I found more interesting, and polymers won. Those meetings proved crucial for helping me to understand where the cutting edge for creating new polymer materials really was at the time. I saw a couple of opportunities. One was to find ways to give biodegradable materials different physical properties and use those materials in high-value applications such as in biomedical devices.

### How did you approach your job search?

I applied for jobs probably even before I was ready for them, and found that it really helped me to hone my research proposals. In 2004, I started applying for academic posts. Rather than looking for jobs with ‘inorganic’ in the advert, I applied for a UK fellowship in nanoscience, and persuaded the university that I had the skills for the job. In 2005, I started my own group at the University of Warwick.

### What are you working on at the moment?

I’m working on degradable polymers. One is a hydrogel material that could one day be combined with adult stem cells to make a scaffold able to regenerate a human spinal disc. Once the cells start to grow, the biological material would take over, leaving nothing synthetic in the body.

### How do you feel about media observations that Royal Society of Chemistry award winners often become Nobel laureates?

I find it quite amusing. It would be lovely if that happened, but I think the press made a bit of a leap. I don’t feel daunted by it because I don’t take it seriously. ■

INTERVIEW BY VIRGINIA GEWIN

# GOOGLE CAR TAKES THE TEST

*The drive of your life.*

BY NORMAN SPINRAD

W hadya mean a ticket? Come on, I'm a cop too, ain't I? Well almost, so I'm only a driving-test officer, but I got the cap and the uniform, don't I? And I wasn't driving this damn thing, now, was I?

Just my job, pal. Not my idea that Google Cars should have to have driving licences to hit the road like you and me. Good or bad idea, it's the law, and we didn't make it, did we? My job is to test the thing, pass or fail, and this sucker didn't, and your job is to write a ticket, or yeah, in this case, a whole padfull, if a driver plays wise guy with the rules of the road.

But you can't ticket me, I wasn't driving, not for a minute. Take a good look, pal. No steering wheel, no brake pedal, no accelerator, no speedometer, no nothing. Yeah, you gotta write up a tonne of tickets, but you gotta ticket Google.

I wasn't exactly against these driverless cars before, in my line of duty, you spend your working day being driven around by so-called human drivers half of which just manage to pass the road test and the other half who don't but usually don't quite get you killed.

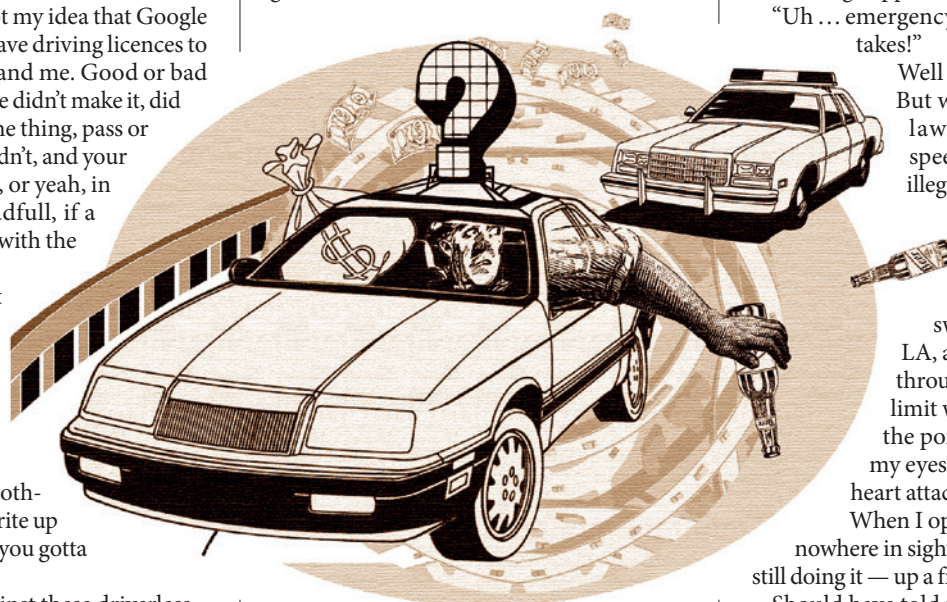
Which is why we got a reputation for looking for ways to fail you, especially towards the end of the shift, I mean, officer, wouldn't you? And seeing as how I see myself as a guardian of the safety of the road from what shouldn't be allowed to barrel along down it, I'm not against making these robots pass the test too.

Okay, I gotta admit I had it in for the thing, especially when it's last on the line at the end of the day, looks like one of those cartoon cars out of that Disney movie, what can I tell you, I got three kids, and I just can't take any more cutesy-poo than I have to, can you?

It's passed the written test or it wouldn't be on the line, how could it not since it's really Google, and Google knows everything.

It works on voice command, so I tell it to open up and it lets me in. I tell it to close, but it won't do that until I buckle in, fair enough. Not every

test car comes equipped with dual steering wheels and peddles, it's unfortunately not mandatory, and you get used to it. But when it gets through to me that this thing doesn't have any manual controls at all I gotta admit it freaks me out a bit. However, it's the end of the shift, and the job is the job, what are you gonna do, so ...



I tell it to pull away from the kerb, which it does, but it immediately loses points for not looking back over its shoulder and hand-signalling, which, having no hands, head or shoulder, it can't, but those are the rules, dumb as they may be in modern times.

Keeps good distance front and back in the straightaways, doesn't try to race through yellow lights, turns right from the right lane and left from the left and all, maintains speed limits exactly like a prissy sissy, which is what you're supposed to do on the test. But keeps losing points for not hand-signalling so I've got plenty enough to fail it with even before the parking test.

So I pick a space between a Toyota and a VW with maybe six inches front and back to squeeze into, this I gotta see ...

"Pull over and park here."

Which, like magic, it does!

Only it cuts off a Hell's Angel on his Harley doing it, and this dude is not amused, starts dismounting with his engine still running, blood in his eyes, and a big monkey wrench in his hand.

"Get us out of here right now!"

It pulls out sideways, but just sits there.

"Go straight!"

It does. But the Harley is following us and the Google Car keeps to the speed limit like a good little citizen, which, as you gotta know, officer, is not a constraint on the Hell's Angel.

"Don't let that bike catch up with us!"

Nothing happens.

"Uh ... emergency override! Do what it takes!"

Well the Google Car does. But within the letter of the law. Doesn't break the speed limit. Doesn't make illegal turns. What it does is find itself a one-way street jammed with crawling early rush-hour traffic, no sweat for Google GPS in LA, and it bobs and weaves through it at the exact speed limit with inches to spare, to the point where I gotta close my eyes to keep from having a heart attack.

When I open them, the Harley is nowhere in sight, but the Google Car is still doing it — up a freeway on-ramp!

Should have told it what? In the middle of a crowded freeway on-ramp? Stop? Turn back? Wait your turn?

What would you have said, wise guy?

Whatsa matter, officer, cat got your tongue?

Okay, okay, I shouldn't have said what I did, and if I hadn't maybe, just maybe, I wouldn't have found myself trapped in a so-called smartcar jiggling and jaggling at the legal speed limit through three lanes of traffic going 10 m.p.h. faster through the desert with a dozen Highway Patrol cars tailing it and trying to figure out what in hell to do about it until its battery finally ran down.

I didn't even realize it wasn't just inside my head when I screamed it. In fact, I think the first part was just in my head ... Google, Schmoogole, these things ...

What did I tell the Google Car out loud? Tell me you wouldn't have said it, officer!

"Drive like a drunken bank robber with a sack full of hundred dollar bills on his way to Las Vegas!" ■

**Norman Spinrad** has now been publishing novels in English for an actual half-century. His latest publication in English is the pamphlet *Raising Hell*. His latest novel has just been published in French as *Police Du Peuple*.

JACEY

➤ NATURE.COM

Follow Futures:

Twitter @NatureFutures

Facebook go.nature.com/mtoodm